Vol. 52 JUNE 1961 No. 3

Journal of Educational Psychology

CONTENTS

Effects of Extended Effort in Creative Problem Solving 117
SIDNEY J. PARNES

Patterns of Self-Conceptualization in High School and College Students 123
WILBUR C. WASHBURN

Manifest Anxiety and Achievement Test Performance 132
HAYNE W. REESE

Creative and Academic Performance among Talented Adolescents 136
JOHN L. HOLLAND

Behaviors during Problem Solving among Children of Low, Average, and High Intelligence 148
HERBERT J. KLAUSMEIER AND LEO J. LOUGHLIN

The Relation between Several Parent Measures and the Child's Early Adjustment to School 153
GENE R. MEDINNUS

On the Use of a Short Vocabulary Test to Measure General Intelligence . 157
JOHN B. MINER

Visual Perception of Binary Patterns by Preschool Children and by School Children 161
DOROTHY W. DYER AND E. RAE HARCUM

Fixed Sequence versus Branching Autoinstructional Methods 166
HARRY F. SILBERMAN, RALPH J. MELARAGNO, JOHN E. COULSON, AND DONALD ESTAVAN

A Re-examination of College Productivity 173
ALEXANDER W. ASTIN

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.
Copyright, 1961 by the American Psychological Association, Inc.

JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 52 June 1961 Number 3

EFFECTS OF EXTENDED EFFORT IN CREATIVE PROBLEM SOLVING[1]

SIDNEY J. PARNES[2]

University of Buffalo

Several studies have demonstrated the possibility of the deliberate development of individual creativity through educational programs (Parnes, 1958, 1960; Taylor, 1959). Two experiments evaluated the creative problem solving course (Parnes, 1959) conducted at the University of Buffalo since 1949 (Meadow & Parnes, 1959; Parnes & Meadow, 1960). In the creative problem solving course a variety of principles and procedures are used to develop creative ability. Two studies have demonstrated the effectiveness of the particular brainstorming principle (Osborn, 1957) of deferred judgment during idea production (Meadow, Parnes, & Reese, 1959; Parnes & Meadow, 1959). Taylor studied the use of the principle among groups vs. individuals (Taylor, Berry, & Block, 1957).

An additional result of the Parnes and Meadow (1959) study indicated a positive correlation between quantity and quality of ideas. The relationship between quantity and quality of ideas has been of great interest in the study of creative problem solving activities. One principle stressed in the course is that copious idea production leads to more creativity in problem solving. Osborn and others have emphasized this need for fluency of ideas. Instructors have reported a prevalence of better ideas at the end of an idea production period than at the beginning (Parnes, 1959). Previous experimenters found increases in remoteness of response and uncommonness of response among succeeding responses on creative tasks (Christensen, Guilford, & Wilson, 1957).

The present study involved two experiments designed to test the principle that extended effort in idea production will lead to an increasing proportion of good ideas with increased production.

Experiment I

The hypothesis was formulated in Experiment I that more good ideas will appear in the last half of a subject's total idea output regarding a creative thinking problem than during the first half. It was also decided to test for any difference between high and low producers in the proportions of their good ideas that fall in the second half.

[1] This research was financed by a grant from the Creative Education Foundation.

[2] The author is indebted to Hayne Reese for his constructive review of this paper, as well as for his assistance with the statistical analyses.

Method

Subjects. Data were available from two groups of subjects. Group 1 consisted of 68 students registered (but uninstructed) in the creative problem solving course at the University of Buffalo. Group 2 was comprised of 78 students enrolled in other undergraduate courses at the university. The data for the two groups were not combined because a cursory examination gave reason to suspect that the groups might differ. (Their averages on the experimental problem were 7.9 and 6.7, respectively.) Accordingly, a statistical treatment was applied which allowed the use of all the data, but took into account this possible difference.

Experimental problem. The Hanger problem used in the present experiment was selected from Part V of the AC Test of Creative Ability. This test is reported to have differentiated groups of creative from noncreative subjects (Harris & Simberg, 1954). Data are available from a previous experiment which indicated that scores on the Hanger problem are positively correlated with other tests which have been designed to measure creative thinking (Meadow & Parnes, 1959).[3] Instructions for the Hanger problem required subjects to list all possible uses for an ordinary wire coat hanger. Time limit for the test was 5 minutes.

Ratings. The test was group administered during regular class periods. The scoring procedure was modified to yield a quantity and *quality* score instead of a quantity and *uniqueness* score, as reported in the previous experiment (Meadow & Parnes, 1959). Each response was scored, by a trained research assistant, as indicating either good or bad quality. The quality score was defined as comprising two dimensions: uniqueness—degree to which the response departed from the hanger's conventional use and value—the degree to which the response was judged to have social, economic, esthetic, or other usefulness.

The scorer was instructed to rate each response on a three-point scale for uniqueness and a three-point scale for value. The response was finally scored as indicating good quality if assigned a combined uniqueness and value score of at least 5. Final quality score used was the total number of "good quality" responses. Any response which duplicated (in essential meaning) responses already given was eliminated from the scoring.

The rater had no knowledge of the purpose of this study; therefore, there was no reason to suspect any bias in ratings of early ideas vs. later ideas. Interrater reliability for ratings of 50 subjects selected at random was .74.

Procedure. The subjects were divided into subgroups on the basis of their total quantities of ideas. Group 1 was divided at its median of 7.9, and Group 2 at its median of 6.7. The subjects with a total number of ideas less than the median were assigned to Subgroup L, and those with a total number of ideas greater than the median were assigned to Subgroup H. To obtain proportional cell frequencies for the analysis of variance, the data of 12 subjects, 6 from each group, were eliminated in a random manner, with the restriction that their total number of ideas was near the median. There remained 62 Group 1 and 72 Group 2 subjects.

Each subject's data were divided into halves on the basis of his total number of ideas. Those subjects with an odd number of ideas were treated as follows: If the middle idea was a good one, it was divided between the halves by adding one-half point to the score for each half.

Results and Discussion

Table 1 presents the mean number of "good" ideas in each half for each group. There were more good ideas in the second half than in the first half, and more good ideas in Subgroup H than in Subgroup L, but there was little difference between Group 1 and Group 2. A Lindquist Type III analysis of variance of these data, (Lindquist, 1956) indicated highly significant differences between Half I and Half II ($F = 32.16$; $df = 1{,}130$; $p < .001$) and between Subgroups L and H ($F = 56.37$; $df = 1{,}130$; $p < .001$). No other differences were statistically reliable.

The results of the experiment indicate that extended effort in producing ideas on a creative thinking problem tends to reward subjects with a greater

[3] Scores on the Hanger problem in the previous experiment correlated with other creative ability tests as follows: Guilford's Unusual Uses, .473; Guilford's Plot Titles High, .452; Guilford's Apparatus, .301; TAT Originality, .520. All but the Apparatus correlation were at the .01 level of significance. The correlation with the Apparatus test was significant at the .05 level.

proportion of good ideas in the second half of their total output. As in a previous experiment (Parnes & Meadow, 1959) a significant relationship was found between quantity and quality of ideas. There was no significant difference between high and low producers in their *proportions* of good ideas in the second half of total production.

The findings support Osborn's theory (1957) that in idea production, quantity leads to quality. The results also seem to be in accordance with Gordon's explanation of "deferment" in the creative process. He describes deferment as "the capacity to discard the glittering immediate in favor of a shadowy but possibly richer future" (Gordon, 1956). The *noncreative* problem solver gets an idea, sees it as a possible solution to his problem, and settles for it without further ado. The *creative* problem solver is not satisfied with his first idea. Like the person who invests money to obtain greater rewards later, the creative person forgoes the immediate reward of applying his first idea, in expectation of a better solution (greater reward) ultimately. A further hypothesis suggested by Osborn's and Gordon's theories is that the *best* idea will come late in the total production period. The author is currently considering experiments to test this particular hypothesis.

Experiment II

Experiment II was designed to determine whether the type of results found in Experiment I with untrained subjects would also occur with subjects trained in the use of the brainstorming principle of deferred judgment. It was also decided to lengthen the time period to 15 minutes instead of 5, in order to see if a trend could be observed towards increasingly greater proportions of good ideas as a subject's total quantity increases.

TABLE 1

Mean Number of Good Ideas among First and Second Halves of All Ideas and among All Ideas

Group	First Half	Second Half	All Ideas
Group 1			
Low	.69	1.11	1.81
High	1.61	2.48	4.10
Mean	1.15	1.80	2.95
Group 2			
Low	.51	1.24	1.75
High	1.47	2.42	3.89
Mean	.99	1.83	2.82
Combined			
Low	.60	1.18	1.78
High	1.54	2.45	3.99
All Subjects Mean	1.07	1.81	2.88

Note.—Low refers to the group whose total number of ideas was less than the median number of ideas; High refers to the group whose total number of ideas was greater than the median number of ideas.

Method

Subjects. Forty-two students in two creative problem solving classes at the University of Buffalo served as subjects. The experiment was conducted during the final half of the course.

Experimental problem. The Hanger problem was used as in Experiment I. However, the trained students were very familiar with the brainstorming principle. Therefore, they were urged to follow this principle of deferred judgment during their idea production. Time limit was 15 minutes.

Ratings. The same rating system was used as in Experiment I. However, before the new experiment was conducted, the rater had already been told the purpose of the original study. If he were given the subjects' protocols, his ratings of the ideas might have been influenced this time by the relative position of each idea on the page. Accordingly, the rater was given each subject's list of ideas typed onto new sheets in alphabetical order. Thus he did not know the position of an idea in a subject's total list of ideas. After the rater had scored the ideas on the

typed sheets, clericals transferred these ratings back to the subjects' original lists.

Procedure. The total list of ideas of each subject was then divided into thirds—a first third, a middle third, and a last third in a similar manner as was used to divide the total ideas into halves in Experiment I.

Results and Discussion

The mean numbers of good ideas per third were 4.84, 5.09, and 5.81 for the first, middle, and last thirds, respectively. The analysis of variance of these data indicated that the main effect of thirds was significant ($F = 4.12$; $df = 2,84$; $p < .05$). t tests showed that the mean for the last third was significantly greater than the means for both the first third ($t = 2.75$; $df = 84$; $p < .01$) and the middle third ($t = 2.04$; $df = 84$; $p < .05$). The means for the first and middle thirds did not differ significantly ($t < 1.00$). Thus the findings of Experiment I are supported, and a trend is suggested towards increasingly larger proportions of good ideas with increased quantity. The results seem to provide further interest in the hypothesis that the *best* idea will come late in the total production period.

The data of the first 5 minutes only[4] were examined on the same basis as Experiment I. For these first 5 minutes, a mean of 3.02 good ideas were produced in the first half of total ideas and a mean of 3.26 in the second half. Analysis of variance of these data indicated that these means were not significantly different. It is interesting to note that these means are both almost double to triple the means for the untrained subjects reported in the results section of Experiment I. This supports the evidence of the earlier experiments which demonstrated the effectiveness of the brainstorming principle and of training in its use (Meadow et al., 1959; Parnes & Meadow, 1959).

[4] Subjects had been requested at periodic time intervals during Experiment II to draw a line under their last idea.

Discussion

Examination of the protocols suggested that the untrained subjects of Experiment I seem to start their production with more of the conventional, unoriginal ideas—high in usefulness but low in uniqueness. Toward the end of a 5-minute period, however, these untrained subjects may find themselves exhausted of these ideas, and may begin groping for less obvious ones. This may lead to their greater production, among the later ideas, of ones high in both usefulness *and* uniqueness, the criteria for the "good" ideas. This seems to be consistent with the findings of Christensen et al. (1957), where uncommonness and remoteness of responses increased with succeeding responses. In the protocols of the *trained* subjects of Experiment II, there seem to be many of the unique and useful ideas among the first ones; but within a 15-minute period these trained subjects may also tend to exhaust themselves of ready ideas and find that they too have to "stretch" their imaginations more fully. This may explain why their strongest production of good ideas in the 15-minute period does not appear until the last third of their ideas.

In order further to explore the difference in results between trained and untrained subjects, a sign test analysis was made of scored data on hand from 180 *untrained* subjects of a previous experiment (Meadow & Parnes, 1959), for an entirely different type of problem (Guilford's Plot Titles) with a 3-minute test period. These data indicated no significant difference in number of quality ideas produced in first half (mean of .57) vs. second half (mean of .62) of the idea lists.

Thus we have data suggesting: (*a*)

for *untrained* subjects, significantly more good ideas among *later* ideas in a *5*-minute period, but not significantly more in a *3*-minute stint; (*b*) for *trained* subjects, significantly more good ideas among *later* ideas for a *15*-minute period, but not significantly more for a *5*-minute stint. Theorizing from the four sets of related data, it might be suggested that subjects will get at least as many good ideas in the second half of their total idea output regarding a creative thinking problem as in the first half, but that with increasing production time, the proportion of good ideas in the second half will increase over the first half. It might be further theorized that untrained subjects will show this increasing proportion more rapidly than trained subjects, who have already learned to be more creative in their early ideas. In a sense, both trained and untrained subjects seem to be stimulating their creativity by *extended* effort in idea-finding. For the trained subjects, a 5-minute period (on the type of problem used) does not seem to be long enough to elicit *extended* effort on their parts. (This might be related analogously to the accomplished musician practicing on a beginner's piece.) However, for untrained subjects 5 minutes *did* seem to provide enough time to call for *extended* effort. Much further experimentation is needed with both trained and untrained subjects, given a variety of time limits and types of problems, in order to verify the theory suggested.

Summary

Experiment I was designed to test the hypothesis that more good ideas will appear in the last half of a subject's total idea output regarding a problem than during the first half. Each subject was given a creative thinking problem which involved production of ideas for a 5-minute period. The resultant ideas were evaluated by a trained rater. Number of good ideas was tallied for the first half of each subject's total idea list, and for the second half of his list.

The findings demonstrated significantly more good ideas to appear in the last half than in the first half of the idea lists. A significant relationship was also found between total quantity and total quality scores.

Experiment II was designed to determine whether the type of results found in Experiment I with untrained subjects would also occur with subjects trained in the use of the brainstorming principle of deferred judgment. It was also decided to lengthen the time period to 15 minutes instead of 5, in order to see if a trend could be observed towards *increasingly* greater proportions of good ideas as a subject's total quantity increases. Each subject was given 15 minutes to produce ideas regarding the same creative thinking problem as in Experiment I. Resultant ideas were evaluated as in the first experiment. Number of good ideas was tallied for each third of each subject's total idea list—first third, middle third, and last third.

The findings showed significantly more good ideas to appear in the final third of the subjects' idea lists. Although there was not a significant difference between first and second "thirds," a trend is suggested towards increasingly larger proportions of good ideas with increased quantity.

The results are interpreted to indicate that extended effort in producing ideas on a creative thinking problem tends to reward problem solvers with a greater proportion of good ideas among the later ideas on their lists.

REFERENCES

Christensen, P. R., Guilford, J. P., & Wilson, R. C. Relations of creative re-

sponses to working time and instructions. *J. exp. Psychol.*, 1957, **53**, 82–88.

GORDON, W. J. J. Operational approach to creativity. *Harv. bus. Rev.*, 1956, Nov.–Dec., 41–51.

HARRIS, R. R., & SIMBERG, A. L. *AC Test of Creative Ability: Examiner's manual.* Detroit: General Motors Corp., AC Spark Plug Division, 1954.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

MEADOW, A., & PARNES, S. J. Evaluation of training in creative problem solving. *J. appl. Psychol.*, 1959, **43**, 189–194.

MEADOW, A., PARNES, S. J., & REESE, H. Influence of instructions and problem sequence on a creative problem solving test. *J. appl. Psychol.*, 1959, **43**, 413–416.

OSBORN, A. F. *Applied imagination.* New York: Scribner, 1957.

PARNES, S. J. (Ed.) *Compendium of research on creative imagination.* Buffalo, N.Y.: Creative Education Foundation, 1958.

PARNES, S. J. *Instructor's manual for semester courses in creative problem-solving.* Buffalo, N.Y.: Creative Education Foundation, 1959.

PARNES, S. J. (Ed.) *Compendium No. 2 of research on creative imagination.* Buffalo, N.Y.: Creative Education Foundation, 1960.

PARNES, S. J., & MEADOW, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *J. educ. Psychol.*, 1959, **50**, 171-176.

PARNES, S. J., & MEADOW, A. Evaluation of persistence of effects produced by a creative problem-solving course. *Psychol. Rep.*, 1960, **7**, 357–361.

TAYLOR, C. W. (Ed.) *The third University of Utah research conference on the identification of creative scientific talent.* Salt Lake City, Utah: Univer. Utah Press, 1959.

TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking. *Admin. sci. Quart.*, 1958, **3**, 23–47.

(Received October 3, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 123-131

PATTERNS OF SELF-CONCEPTUALIZATION IN HIGH SCHOOL AND COLLEGE STUDENTS

WILBUR C. WASHBURN
University of California

Self-concepts have been related to aspects of child development in the theoretical orientations of various psychologists—Freud, Erikson, Sarbin, Horney, and Fromm. Although the specific term "self-concept" has not always been used, the "character traits" described by these writers correspond to what is now frequently called a self-concept. For purposes of this paper, a self-concept is regarded as a relatively enduring aspect of personality—a motive, attitude, or value—by means of which the individual relates himself to his social environment. The term applies to the person's sets of ideas about himself in relation to other persons.

Procedure

The problem was to find patterns of self-concepts useful in describing adolescents and young adults and to integrate the various theories that formulate self-concepts in terms of child development. Since many existing sets of postulated self-concepts were originally inferred on the basis of very early developmental periods, they may not be adequately descriptive of more mature levels of personality. The procedure, then, was to translate the hypothetical patterns of self-concepts offered by a variety of theorists into test items, to administer the test to groups of adolescent and young adult subjects, and to revise the list of hypothetical patterns of self-conceptualization on the basis of intercorrelation of test items. The resulting sets of self-concepts were then ranked on the basis of maturity ratings provided by groups of graduate students specializing in developmental psychology.

Since the results of any statistical analysis of test items depend in part on the nature of the items selected for analysis, an effort to minimize this limitation was made by including items based on a variety of theoretical viewpoints. Table 1 indicates the postulated patterns of self-concepts for which items were devised and lists a sample item for each category. Items were based on the following theoretical formulations: (*a*) the empirical selves hypothesized by Sarbin (1952), (*b*) the psycho-social stages stated by Erikson (1960), (*c*) the character types described by Freud (1933), (*d*) the personality patterns postulated by Fromm (1947), and (*e*) the interpersonal attitudes developed by Horney (1945).

Construction of the Testing Instrument

The test originated with a cluster analysis, using the technique described by Tryon (1939), applied to Sarbin's five empirical selves. The author constructed 135 items representing self-concepts postulated by Sarbin. Five graduate students specializing in personality theory were asked to classify each item into the category it was supposed to measure. All five judges correctly classified 110 items, which then were administered to a group of 100 high school students.

Four patterns of self-concepts emerged as a result of the cluster analysis. Sarbin's first three empirical selves were highly intercorrelated and

TABLE 1

THE POSTULATED PATTERNS OF SELF-CONCEPTS ON WHICH TEST ITEMS WERE BASED

Personality Theorist	Hypothesized Self-Concept	Example of Item[a]
Erikson	Infancy (Trust vs. Mistrust)	Most persons can be trusted.
	Early Childhood (Autonomy vs. Shame and Doubt)	I prefer to do things for myself.
	Play Age (Initiative vs. Guilt)	I take the lead in getting friends together.
	School Age (Industry vs. Inferiority)	I really enjoy working.
	Adolescence (Identity vs. Identity Diffusion)	I usually stick to my own ideas even when friends disagree.
	Young Adulthood (Intimacy vs. Isolation)	When feeling blue I try to find someone who is cheerful.
	Adulthood (Generativity vs. Self-Absorption)	I wish I could discover something that would benefit future generations.
	Senescence (Integrity vs. Disgust)	I try to find my faults and correct them.
Freud	Oral Character	I look on the bright side of things.
	Anal Character	I tend to save my money.
	Genital Character	I love humanity in general.
Fromm	Receptive Character	I prefer letting others make plans.
	Exploitative Character	I want to win when pitted against others.
	Hoarding Character	If buying a car, I would be most interested in its economy.
	Marketing Character	It is important to always please others.
	Productive Character	Friends say I am creative and original.
Horney	Compliant (Moving Toward)	I enjoy just being with people.
	Detached (Moving Away)	I prefer being alone.
	Aggressive (Moving Against)	I often object to another's ideas.
Sarbin	Somatic Self	My body has greatly influenced me.
	Receptor-Effector Self	I can't wait to eat when hungry.
	Primitive-Construed Self	I let off steam when angry.
	Introjecting-Extrojecting Self	I usually do the expected thing to avoid disapproval of others.
	Social Self	If I cheated on a test, I would feel so guilty I couldn't stand it.

[a] Items were selected randomly for illustration.

were labeled the Somatic-Primitive Self. Items from Sarbin's fifth empirical self had a high negative correlation with the first cluster and were called the Inner-Controlled Self. Sarbin's fourth empirical self split into two sets of items. One cluster was labeled the Submissive-Dependent Self, while the other was tentatively named the Socially Integrated Self. These labels were assigned each cluster by considering apparent logical relationships among items.

Items representing a wider variety of theories were then added. The author constructed 250 items based on the theories of Erikson, Freud, Fromm, and Horney. The judges were again asked to classify each new item into the category or categories it presumably measured, and the 225 items on which there was 100% agreement were administered to a second group of 100 high school students. A biserial coefficient of correlation was computed to show the relationship of each item to the total score of the four existing clusters. Items were assigned to the cluster with which they had the highest correlation.

All items representing Horney's compliant orientation and many items describing Fromm's receptive and marketing characters were added to the test measuring the Submissive-Dependent Self. A large number of items were added to the test measuring the Inner-Controlled Self from those representing Freud's anal character, Fromm's hoarding type, and the problems listed by Erikson under school age, adolescence, adulthood, and senescence. The cluster labeled Socially Integrated Self absorbed many items representing Freud's oral and genital characters, Fromm's productive character, and the problems stated by Erikson under adulthood and senescence. Inspection of the items added to this cluster suggested a similarity to the self-actualizing personality described by Maslow (1954), and the cluster was renamed the Integrative-Actualizing Self. A set of items was found to have a high negative correlation with the Submissive-Dependent Self and was labeled the Outer-Controlling Self. This new pattern of self-concepts was composed largely of items representing Horney's aggressive orientation, although it also contained items measuring the problems described by Erikson as typical of early childhood, the play age and the school age. Another group of items had a high negative correlation with the Integrative-Actualizing Self and was named the Detached-Independent Self. It was composed of items measuring Horney's detached orientation and the problems outlined by Erikson with regard to the periods of infancy, early childhood, and young adulthood.

Table 2 indicates relatively high negative correlations between certain of the six resulting patterns of self-concepts. Each pair of clusters with high negative correlations described contrasting self-concepts. Hence forced-choice items were developed in which items from negatively correlated clusters were combined as alternative choices in a single question. Subtest 1 measured the Somatic-Primitive Self and the Inner-Controlled Self. Subtest 2 measured the Submissive-Dependent Self and the Outer-Controlling Self. Subtest 3 measured the Detached-Independent Self and the Integrative-Actualizing Self. Sample forced-choice items from each subtest are presented in Table 3.

The use of forced-choice items offered several advantages. It resulted in a 50% reduction in the number of answers to be scored on each answer sheet. Once a score was obtained for a given cluster, the score for its negatively correlated counterpart could be computed simply by subtracting the first score from the maximum attainable score. Anastasi (1954) points out that this testing technique provides changing alternatives and thus avoids

TABLE 2

INTERCORRELATIONS AMONG SELF-PATTERNS

	Submissive-Dependent Self	Detached-Independent Self	Outer-Controlling Self	Inner-Controlled Self	Integrative-Actualizing Self
Somatic-Primitive Self	−.17	.10	.21	−.83	−.12
Submissive-Dependent Self		.10	−.86	.18	−.08
Detached-Independent Self			−.09	−.01	−.78
Outer-Controlling Self				−.19	.11
Inner-Controlled Self					.13
Integrative-Actualizing Self					

TABLE 3
SAMPLE ITEMS FROM THE THREE SUBTESTS

Subtest	Examples of Test Items[a]	
1		
(A) Inner-Controlled Self	(A) If I won $1,000, I would be more likely to save most of it for future needs.	(B) If I won $1,000, I would be more likely to spend most of it now on things I need to have a good time.
(B) Somatic-Primitive Self	(A) I would prefer having parents, teachers, or employers think well of me.	(B) I would prefer just taking it easy and enjoying life.
	(A) My family has had a greater influence on me.	(B) My body and its feelings have had a greater influence on me.
	(A) I say little when angry.	(B) I let off steam when angry.
	(A) I am sure most policemen are helpful.	(B) It is likely too many policemen try to show their authority.
2	(A) I like giving directions to others.	(B) I prefer following directions given by others.
(A) Outer-Controlling Self	(A) I want to be considered strong by others.	(B) I prefer being thought of as kind and sympathetic.
(B) Submissive-Dependent Self	(A) I would like to be thought of as an important person in my community.	(B) I wouldn't care to be considered an important person.
	(A) I often raise objections to another's ideas.	(B) I usually accept the ideas of others without being critical.
	(A) My former teachers would be more likely to call me a leader.	(B) My former teachers would be more likely to call me a follower.
3		
(A) Integrative-Actualizing Self	(A) I often enjoy change and rather look forward to trying new things and meeting new people.	(B) I dislike change and tend to feel uncomfortable in new situations.
(B) Detached-Independent Self	(A) One of my best features is that I like people.	(B) One of my best features is that I don't let feelings influence my judgment.
	(A) I realize human nature isn't perfect, but I can accept it and even like it as it is.	(B) I find it hard to accept human nature as it is.
	(A) I am left with a feeling of strength upon seeing or hearing something truly beautiful.	(B) I am not too impressed by art or music, although these things seem beautiful to others.
	(A) I am usually optimistic and enthusiastic and seldom lonely and depressed.	(B) There are quite a few times when I feel alone and afraid and just about the whole world seems unfriendly.

[a] Items were selected randomly for illustration.

the tendency toward a response set characteristic of constant-alternative, true-false items.

Forced-choice alternatives were selected from items which seemed to logically fit together. However, items chosen by over 90% or fewer than 10% of the total group were eliminated, since they seemed to reflect extremes of social approval and disapproval. New forced-choice items, 50 for each subtest, were added to existing items and the revised test administered to a third group of 100 high school students. A biserial coefficient of correlation was computed showing the relationship of each item to the total score. Items correlating less than .20 with the total subtest were eliminated. The new forced-choice form was then administered to a group of 300 undergraduate students at the University of California, each item correlated with the total score, and additional items eliminated.

Validity of the subtests was appraised by asking high school counselors to rate students for each subtest on the basis of an adjective checklist describing characteristics appropriate for the subtest. For example, descriptive terms applicable to the first subtest included the following words: conscientious, controlled, conforming. Subjects with ratings in the upper and lower 27% of the group based on total number of adjectives checked for each subtest were then compared for mean scores on the subtest and t ratios computed. The differences between means were all significant at less than the .01 level of confidence. Similar results were obtained in a second validation study. Corrected split-half reliability coefficients were above .80 for groups from the tenth grade to college undergraduates.[1]

[1] A copy of the testing instrument has been deposited with the American Documentation Institute. Order Document No. 6612 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Five graduate students at the University of California were asked to rank each cluster in terms of relative maturity as inferred on the basis of their knowledge of developmental psychology. Each judge had completed at least two advanced undergraduate courses and graduate research in this area. A second group of five graduate students with similar qualifications but from a different institution also ranked the six patterns of self-concepts. While both groups agreed on the rank order, agreement was less consistent for the relative position of the Detached-Independent Self and the Outer-Controlling Self than for the other levels of self-conceptualization. There was 100% agreement on the Somatic-Primitive Self; 90% each on the Dependent-Submissive Self, the Inner-Controlled Self, and the Integrative-Actualizing Self; 70% on the Detached-Independent Self; and 60% on the Outer-Controlling Self.

The six resulting self-patterns, listed from the one rated least to the one ranked most mature, may be described as follows: (*a*) the Somatic-Primitive Self involves inability to postpone need satisfaction and failure to identify with authority figures, (*b*) the Submissive-Dependent Self involves avoidance of disapproval by submitting to others, (*c*) the Detached-Independent Self involves avoidance of interpersonal relations and situations likely to arouse emotion, (*d*) the Outer-Controlling Self involves achievement of social status through manipulation of the external world in socially approved ways, (*e*) the Inner-

Controlled Self involves development of inner controls through internalization of external social norms, (*f*) the Integrative-Actualizing Self involves acceptance of oneself and others and a tendency toward increased creativity and productivity.

Results

A Comparison of High School and College Groups

The final form of the test was utilized with a sample of high school and college students. The high school students were from a large urban high school in California and were randomly selected within categories to constitute a stratified sample proportionate to the college group regarding racial and socioeconomic classification. The college students consisted of undergraduates attending the University of California at Berkeley, and the sample was selected to be representative of this campus with regard to fields of study. Separate results are reported for males and females in both groups and for college preparatory and noncollege preparatory high school students. The college preparatory group consisted of students taking courses which will enable them to attend a private or state college, a private or state university, or a junior college. The college sample constitutes a group selected from a sample similar to the college preparatory group on the basis of academic achievement in high school. Further details regarding these groups are presented in Table 4.

Means for high school and college groups are indicated in Table 5. Separate means are not given for each age group within the high school or college groups, since no statistically significant differences were found with regard to such age levels. Since each subtest measures two negatively correlated self-levels, one rated as relatively more mature than the other, the means presented in Table 5 are for the self-pattern with the higher maturity ranking. For example, the

TABLE 4

Composition of Groups Used in the Study

Characteristics	College Sample		High School: Noncollege Preparatory		High School: College Preparatory	
Males	66		40		100	
Females	49		45		135	
Median Age (Males)	22 yrs.-1 mo.		16 yrs.-11 mo.		16 yrs.-10 mo.	
Median Age (Females)	21 yrs.-10 mo.		16 yrs.-11 mo.		16 yrs.-9 mo.	
Caucasian	99	86%	73	86%	202	86%
Negro	9	8%	7	8%	19	8%
Other	7	6%	5	6%	14	6%
Upper Socioeconomic Status[a]	32	28%	24	28%	66	28%
Middle Socioeconomic Status[a]	60	52%	44	52%	122	52%
Lower Socioeconomic Status[a]	23	20%	17	20%	47	20%
First Year	15		0		0	
Second Year	61		0		0	
Third Year	20		30		92	
Fourth Year	19		55		143	

[a] Based on father's occupation.

TABLE 5

A Comparison of High School and College Groups

	N	Subtest 1		Subtest 2		Subtest 3	
		M	*SD*	*M*	*SD*	*M*	*SD*
Males							
High School, Noncollege	40	70.1	16.3	50.4	11.6	70.0	15.1
High School, College Preparatory	100	78.2	25.8	58.4	13.9	74.2	20.7
College	66	79.3	16.2	66.7	11.9	76.3	16.6
t (High School, Noncollege vs. High School, College Preparatory)		1.81		3.18**		1.15	
t (High School, College Preparatory vs. College)		.25		3.44**		.57	
Females							
High School, Noncollege	45	71.5	20.4	41.0	15.1	77.5	14.3
High Schoool, College Preparatory	135	79.0	19.1	50.1	14.5	86.6	16.0
College	49	78.0	16.7	55.9	14.4	71.5	19.2
t (High School, Noncollege vs. High School, College Preparatory)		2.46		3.65**		2.27	
t (High School, College Preparatory vs. College)		.62		2.60		4.20**	
Males-Females Compared							
t (High School, Noncollege Females vs. High School, Noncollege Males)		.33		3.06**		2.35	
t (High School, College Preparatory Females vs. High School, College Preparatory Males)		.58		4.39**		3.91**	
t (College Females vs. College Males)		.31		4.38**		1.42	

** Significant at the .01 level of confidence or less.

means given under Subtest 1 are for the Inner-Controlled Self, and the means for the Somatic-Primitive Self can be determined by subtracting the stated means from the maximum possible score of 124 points. Means presented under Subtest 2 apply to the Outer-Controlling Self, and means for the Submissive-Dependent Self can be determined by subtracting the stated means from 103, the highest attainable score for this subtest. Means listed for Subtest 3 represent the Integrative-Actualizing Self, and means for the Detached-Independent Self may be computed by subtracting the stated means from the maximum possible score of 120 points.

Although high school boys preparing for college tended to score higher than those not preparing for college on all three self-levels rated most mature, the difference was significant at less than the .01 level only for the Outer-Controlling Self. Although college males scored higher than high school college preparatory males on all three mature levels of self-conceptualization, differences were slight and not significant at an acceptable level ex-

cept for the Outer-Controlling Self, which involved a difference significant at less than the .01 level.

High school girls preparing for college scored higher than noncollege preparatory females on the three self-levels with highest maturity rankings. Differences were significant at less than the .02 level for the Inner-Controlled Self, at less than the .01 level for the Outer-Controlling Self, and at less than the .05 level for the Integrative-Actualizing Self. Means representing the Outer-Controlling Self were higher for college women than college preparatory high school girls, and the difference was significant at about the .01 level. However, female college students obtained a lower mean than college preparatory high school girls on the test measuring the Inner-Controlled Self, although the difference was not statistically significant at an acceptable level. The college females also had a lower mean than college preparatory females on the test measuring the Integrative-Actualizing Self, and the difference was large and significant at less than the .01 level.

Both groups of high school females scored higher than corresponding groups of males on the test measuring the Inner-Controlled Self, while males scored slightly higher than females on this test at the college level. However, these differences were neither large nor statistically significant. All male groups scored higher than corresponding female groups on the test measuring the Outer-Controlling Self, and all differences were relatively large and significant at less than the .01 level. High school girls had higher means than high school boys for the Integrative-Actualizing Self. The difference between means on this test was significant at almost the .02 level for noncollege preparatory students and at less than the .01 level for college preparatory students. The trend was reversed at the college level, and college men tended to score higher than college women on this test, although the difference was not statistically significant.

Discussion

It is to be expected that college students would tend to score higher than high school students on tests measuring mature self-levels as a result both of college experience and selection of more mature students for admission to the university. This hypothesis tends to be confirmed for males with regard to all three more mature self-levels, although differences are large and significant at less than the .01 level only for the Outer-Controlling Self. Further, females do not seem to follow this trend in relation to the Inner-Controlled Self and the Integrative-Actualizing Self. However, these relatively low scores of University of California females indicate the female group is more highly select than the college males used in this study. Statistics for the high school district show that about 1.5 times more females than males from this area enter private colleges rather than attend public universities. The emphasis on academic achievement as an entrance requirement at the university possibly results in selection of more competitive individuals and would account in part for the relatively high scores of both male and female college groups on the test measuring the Outer-Controlling Self. The tendency for males consistently to score higher than females on the test measuring the Outer-Controlling Self is to be expected in view of child training practices in our society which from early childhood permit males greater freedom than females in exploring and manipulating their en-

vironments. Since females mature both physically and socially at an earlier age than males, it is not surprising to find that high school females tend to score higher than males on the two self-patterns with the highest maturity rankings, while the trend is reversed and males score higher than females at the university level.

The findings for high school groups indicate that college preparatory students tend to have more mature self-concepts than noncollege preparatory students. The large and statistically significant differences obtained between college preparatory and noncollege preparatory students on the test measuring the Outer-Controlling Self suggest that college preparatory students show greater drive to attain status long before attending college. If subsequent research confirms differences in self-concepts related to college attendance and preparation for college, it becomes important to consider personality factors along with academic achievement as a basis for predicting future success among college applicants.

Since the results are likely to vary with the theoretical viewpoints represented in the test items analyzed, the self-patterns described in this study should not be regarded as a final list. Further, the six self-levels should not be interpreted as rigid categories into which all persons may be classified on an all-or-nothing basis. Every subject checked at least some items for every one of the six self-patterns. Hence it seems safe to assume that the more immature levels continue into at least early adulthood to some extent in all individuals.

Summary

Items measuring self-concepts were constructed representing the theories of Sarbin, Freud, Erikson, Horney, and Fromm. A cluster analysis yielded six self-patterns, which, ranked from least to most mature on the basis of ratings by graduate students in psychology, were: (*a*) the Somatic-Primitive Self, (*b*) the Submissive-Dependent Self, (*c*) the Detached-Independent Self, (*d*) the Outer-Controlling Self, (*e*) the Inner-Controlled Self, and (*f*) the Integrative-Actualizating Self. Significant differences in means on tests measuring various self-levels were found for contrasting groups of males and females, college preparatory and noncollege preparatory high school students, and college and high school students.

REFERENCES

Anastasi, A. *Psychological testing.* New York: Macmillan, 1954.

Erikson, E. H. Youth and the life cycle. *Children,* 1960, **7**(2), 43–49.

Freud, S. *New introductory lectures on psychoanalysis.* New York: Norton, 1933.

Fromm, E. *Man for himself.* New York: Rinehart, 1947.

Horney, K. *Our inner conflicts.* New York: Norton, 1945.

Maslow, A. H. *Motivation and personality.* New York: Harper, 1954.

Sarbin, T. R. Contributions to role-taking theory: III. A preface to a psychological analysis of the self. *Psychol. Rev.,* 1952, **59,** 11–22.

Tryon, R. C. *Cluster analysis.* Ann Arbor: Edwards, 1939.

(Received July 13, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 132-135

MANIFEST ANXIETY AND ACHIEVEMENT TEST PERFORMANCE

HAYNE W. REESE

University of Buffalo

According to Spence's (1956) theory the effect of motivational or drive level on performance depends on the complexity of the task. In a "simple" task, in which the correct response tendency is stronger than competing response tendencies, subjects with higher drive perform at higher levels than subjects with lower drive; but in a "complex" task, in which the correct response tendency is weaker than competing response tendencies, they perform at lower levels than subjects with low drive. These effects depend on the energizing property of drive. Castaneda (1956) has suggested that the inverse relationship between drive level and performance may also be obtained in simple tasks, such as a reaction time situation, in which the correct response is mediated by implicit verbal responses. Theoretically, a drive produces "drive stimuli" which may be conditioned to responses. If these drive-produced responses interfere with the implicit verbal responses which mediate the instrumental act, the task becomes a complex one and there will be more interference under high drive, since the drive-produced responses have a greater probability of occurrence, than under low drive.

The Children's Manifest Anxiety Scale (CMAS) was designed by Castaneda, McCandless, and Palermo (1956) as a measure of drive in children. Relationships have been demonstrated between CMAS scores and school achievement (McCandless & Castaneda, 1956) and performance on simple and complex learning tasks (Castaneda, Palermo, & McCandless, 1956; Palermo, Castaneda, & McCandless, 1956) which indicate that children with high CMAS scores are superior to low scoring subjects on simple tasks and inferior on complex tasks, in line with the interpretation of CMAS scores as indicants of drive level.

It has been suggested that the inverse relationship between CMAS scores and performance on complex tasks may result from the negative correlation between CMAS scores and intelligence which has been found (Jafner & Kaplan, 1959; McCandless & Castaneda, 1956; Phillips, King, & McGuire, 1959). Phillips et al. (1959) reported that when the effects of intelligence were partialed out, there was no relationship between CMAS scores and performance on a battery of psychometric tests differing in "complexity."

The present study was designed to provide additional evidence regarding the relationships among manifest anxiety, intelligence, and performance. An arithmetic test was given under a time limit, and it was assumed that for fourth and sixth grade school children the test would be more nearly a speed test than a power test, i.e., that it would be a simple task in which speed was emphasized (as in the reaction time situation). It was further assumed that the motor responses (writing the answers) would be verbally mediated, and that manifest anxiety would produce responses which would interfere with the mediators (i.e., the symbolic processes). Therefore, the inverse relationship be-

tween manifest anxiety and performance was predicted.

Method

Subjects. The subjects were 539 fourth and sixth grade school children from the Williamsville, New York, school district.[1] The data of an additional 85 subjects were eliminated: 54 subjects missed one of the tests, and IQs were not available for 31 subjects.

Tests and Procedure. The subjects were given the CMAS and one week later an arithmetic test consisting of 40 addition problems, each with four three-digit numbers. The arithmetic test had a 13-minute time limit. Both tests were administered by the teachers to their regular classes, using standard instructions. Scores on the Primary Mental Abilities Test, which is routinely administered in the third and sixth grades, were obtained from school records.

Results

Table 1 presents mean scores on the variables, and provides further information about the sample and the difficulty level of the achievement test. (Difficulty level is not necessarily related to "complexity" of the test, since the correct response may be dominant in the response hierarchy but have a low probability of occurrence and therefore a long latency. Under these conditions low scores would result when the time limit is relatively short.)

Correlations among the variables are presented in Table 2.[2] Manifest anxiety was negatively correlated with the number of correct answers on the achievement test, though the correlation was not significant for fourth grade males. To provide further information about the relationship between manifest anxiety and achievement, the subjects of each sex in each grade were divided into six subgroups on the basis of CMAS scores, and their achievement test scores were compared. The main effect of Anxiety Levels was significant ($F = 6.73$; $df = 5,464$; $p < .001$), but this factor did not interact significantly with grade or sex, indicating that there was a significant relationship between manifest

TABLE 1
Mean IQ, CMAS Scores, and Arithmetic Test Scores

Measure	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
N	139	144	118	138
IQ	112.4	108.6	112.3	108.7
Anxiety Score	15.8	14.4	14.9	13.3
Lie Score	5.0	4.3	3.4	3.5
Number Right on Arithmetic Test	11.8	10.9	21.9	20.6
Number Wrong on Arithmetic Test	6.2	6.8	4.3	4.2

[1] The writer is indebted to W. E. Keller, Supervising Principal, Williamsville Central Schools; L. R. Cofran, Principal, Academy Elementary School; R. A. Wilson, Principal, Dodge Elementary School; K. L. Merrick, Principal, Forest Elementary School; and E. Szado, Principal, Maple Elementary School, and their teaching staffs, for their fine cooperation in providing facilities and for their assistance in many other ways. The achievement test was constructed, and the instructions standarized, by P. R. Cutting, Buffalo, New York, whose assistance is gratefully acknowledged.

[2] Castaneda (1956) suggested that there may be two kinds of anxiety, "verbal" and "nonverbal" or "somatic." The former should produce responses which interfere with mediators in the present type of task, but the latter should not. Verbal and somatic subscales were constructed from the CMAS, and it was predicted that the correlations with achievement would be negative for the former and positive for the latter, since subjects with high drive should be superior to subjects with low drive unless the drive produces responses which interfere with the mediators. The average correlations with achievement were −.20 for verbal anxiety and −.15 for somatic anxiety. The failure to confirm the prediction may reflect a lack of validity of the subscales or a flaw in the theoretical analysis.

TABLE 2

PEARSON PRODUCT-MOMENT CORRELATIONS AMONG CMAS SCORES, IQ, AND NUMBER RIGHT ON ACHIEVEMENT TEST

Variables Correlated	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
IQ × R	.46***	.67***	.58***	.45***
Anxiety × R	−.33***	−.15	−.22**	−.19*
Lie × R	−.11	−.22***	−.33***	−.10
Anxiety × IQ	−.23***	−.01	−.18*	.03
Lie × IQ	−.14	−.35***	−.32***	−.17*
Anxiety × Lie	.00	−.24***	−.05	−.23***

Note.—"R" refers to Number Right on the achievement test.

* $p < .05$.

** $p < .025$.

*** $p < .01$.

TABLE 3

PARTIAL CORRELATIONS BETWEEN NUMBER RIGHT AND MANIFEST ANXIETY, WITH IQ PARTIALED OUT AND WITH LIE SCORE PARTIALED OUT

Variable Partialed Out	Group			
	Fourth Grade		Sixth Grade	
	Females	Males	Females	Males
IQ	−.26	−.20	−.14	−.23
Lie Score	−.33	−.22	−.25	−.22

anxiety and performance on the achievement test, but the trends were not significantly different in the different sexes or grades. The eta for the combined sexes and grades was about .21, which was about the same as the mean of the Pearson correlation coefficients between anxiety and achievement (cf. Table 2). Therefore, it was concluded that the relationship between anxiety and achievement is negative and linear.

Examination of Table 2 shows that manifest anxiety was more strongly correlated with achievement test scores than with IQ. Table 3 presents partial correlations between anxiety and achievement, with IQ partialed out, and with lie score partialed out. Partialing out IQ tended to reduce the correlations for females by about .07, and to increase the correlations for males by about .04. Partialing out the lie score increased the correlations an average of about .03.

The multiple correlations for predicting performance from anxiety and IQ were about .04 higher than the correlations for IQ alone; and the combination of anxiety and lie score improved prediction, compared with anxiety alone, by an average of about .07. The use of anxiety, IQ, and lie score improved prediction, compared with IQ alone, by an average of about .04, and compared with anxiety and IQ, by only about .01.

The lie score was significantly negatively correlated with achievement in fourth grade males and sixth grade females. Partialing out IQ reduced the correlations to .03 and −.19, respectively.

DISCUSSION

The data indicate that the relationship between manifest anxiety and performance on a task requiring mediation is monotonic and negative, but that although statistically significant, it is not a strong relationship. Partialing out IQ had little effect on the correlations, but on the other hand the combination of manifest anxiety with IQ did not appreciably increase the accuracy of predicting performance over that for prediction on the basis of IQ alone. Therefore, it must be concluded that although the relationship between manifest anxiety and performance is not a result of the relationship between anxiety and intelligence, manifest anxiety has relatively little influence on scores on an

achievement test of the type used in the present study.

The lie scale does not seem to serve any useful purpose as far as prediction of performance on the present type of task is concerned, though it does appear to measure a tendency to falsify answers, at least in males.

Summary

Fourth and sixth grade children were given the Children's Manifest Anxiety Scale and an arithmetic achievement test, consisting of 40 addition problems, which was given with a time limit. It was found that manifest anxiety was inversely and monotonically related to the number of correct responses on the achievement test. Partialing out IQ had little effect on the correlations between manifest anxiety and achievement, but predictions of achievement were not appreciably improved by the combination of manifest anxiety with IQ.

REFERENCES

Castaneda, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225–228.

Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, **27**, 317–326.

Castaneda, A., Palermo, D. S., & McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, **27**, 327–332.

Hafner, A. J., & Kaplan, A. M. Children's manifest anxiety and intelligence. *Child Develpm.*, 1959, **30**, 269–271.

McCandless, B. R., & Castaneda, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, **27**, 379–382.

Palermo, D. S., Castaneda, A., & McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, **27**, 333–337.

Phillips, B. N., King, F. J., & McGuire, C. Studies on anxiety: I. Anxiety and performance on psychometric tests varying in complexity. *Child Develpm.*, 1959, **30**, 253–259.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received July 25, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 136–147

CREATIVE AND ACADEMIC PERFORMANCE AMONG TALENTED ADOLESCENTS[1]

JOHN L. HOLLAND[2]

National Merit Scholarship Corporation, Evanston, Illinois

The need to understand the nature of academic achievement and creative behavior is pressing; such knowledge is vital for the proper development of educational practice, the administration of scholarship and fellowship programs, and the fostering of talented behavior generally. Unless attempts to encourage the development of creative talent rest on valid assumptions about the nature of academic and creative behavior, such efforts will have little useful influence.

The present study is one of several investigations planned to explore the nature of academic achievement and creative behavior in adolescents and young adults (Holland, 1959a, 1959b, 1960a). The earlier studies were concerned primarily with academic performance. This study tests a large number of hypotheses about variables which are often assumed to be associated with both academic and creative achievement—aptitude scores, self and teacher ratings, originality measures, parental attitudes and values, vocational interests and aspirations, and background information. Most of the variables were used to test a limited number of hypotheses; a few were included for exploratory purposes. The hypotheses about the nature of the academic achiever were derived in part from an extensive literature which suggests that the student who gets good grades in high school is bright, persistent, conforming, self-controlled, responsible, serious, and rated high by his teachers (d'Heurle, Mellinger, & Haggard, 1959; Gough, 1953; Holland, 1959a, 1959b). Drews and Teahan (1957) indicate that the academic achiever tends to have a somewhat authoritarian mother. Several recent reports (Holland, 1960a; MacKinnon, 1959) imply also that good grades in high school and college may be either unrelated to, or negatively correlated with, potential for creative performance. Taken together, these findings have somewhat consistent implications about the academic achiever and his parents' attitudes and values. The hypotheses derived from the literature and tested here are summarized as follows.

1. Outstanding academic achievement—High School Rank (HSR)—will be positively associated with high scores on the Scholastic Aptitude Test (Math and Verbal factors) (SAT-M and SAT-V), and on the Mastery, Deferred Gratification, and Control scales; with high Self-Ratings of Drive to Achieve and Perseverance; with high Teacher Ratings of Maturity, Popularity, and Social Leadership; with Fathers' Values of Good Student; and with mothers' authoritarian attitudes as measured by the Parental Attitude Research Inventory (PARI).

2. Outstanding academic achievement will be associated with low scores on measures of creativity and originality (Differential Reaction Sched-

[1] This study was partially supported by research grants from the National Science Foundation and the Old Dominion Foundation.

[2] I am indebted to the following staff members for their skillful assistance in all phases of this project: Virginia Chalmers, Laura Kent, Donald Thistlethwaite, and Elizabeth van Laer.

ule, Complexity-Simplicity, Independence of Judgment, and Barron Originality scales) and with low self-evaluations.

The hypotheses about the nature of the creative student were drawn principally from the work of Barron (1953a, 1953b), Gough (1957a), and MacKinnon (1959), which depicts the creative person as independent, complex in outlook, curious, self-assured, intellectual, interested in science and art, and generally effective. This characterization implies also a more permissive parental background which encourages independence and the exploration of self and environment. The hypotheses derived from this literature are summarized as follows.

1. Creative performance (winning public competitions which demand creativity) will be positively associated with high scores on measures of creativity and originality (Differential Reaction Schedule, Complexity-Simplicity, Independence of Judgment, and Barron Originality scales), and on the Initiative and Self-Assurance scales; with high Self-Ratings of Originality, Independence, and Perseverance; with Fathers' Values of Curious and Independent; and with mothers' nonauthoritarian attitudes.

2. Creative performance will have no correlation, or negative correlations, with good high school grades; with high Self-Ratings of Popularity; with high Teacher Ratings; and with Fathers' Values of Dependable, Good Student, Happy and Well-Adjusted, and Popular.

Method

Student and Parent Sample

The student and parent samples were obtained from a one-sixth random sample of National Merit Finalists (approximately 9,868 high-scoring students by state from an initial pool of 478,991 high school juniors), who were polled and tested by mail. An 84% return was obtained, but incomplete information reduced the sample to 50%, 649 boys and 345 girls, and their parents. The average aptitude level of these samples on the SAT were: For boys, SAT-V and SAT-M, 658.4 and 697.9 with *SD*s of 56.8 and 62.2, respectively. For girls, the SAT-V and SAT-M means are 659.2 and 635.4 with *SD*s of 60.6 and 70.3.

Academic and Creative Performance

The criterion of academic performance was high school grades (HSR) during the first 3 years of high school. The criteria of creative performance were derived from a checklist of accomplishments assumed to require creative or original behavior. Creative performance is defined as a performance which is accorded public recognition through awards, prizes, or publication, and which may therefore be assumed to have exceptional cultural value. Because of the difficulty in arriving at a generally acceptable definition of "creativity," these criteria should perhaps be regarded as either "notable scientific or artistic performance," although we will refer to the criterion as "creative" performance hereafter to enhance readability. With this definition as a guide, a list of 20 achievements at the high school level was derived by reviewing the secondary school achievements of Finalists from previous years. Items were divided by content into two scales: Creative Science (5 items) and Creative Arts (11 items). (Four of the original 20 items were omitted because they appeared to be inadequate signs of creative behavior.) The items on the two scales are shown below. Students were asked to check those which applied to them.

Creative Science Scale

1. Gave an original paper at a scientific meeting sponsored by a professional society.
2. Won a prize or award in a scientific talent search.
3. Constructed scientific apparatus on own initiative.
4. Invented a patentable device.
5. Had scientific paper published in a science journal.

Item 3 does not meet the criteria of "public recognition" and "unusualness," but in order to lengthen the scale, it was in-

cluded as a low-level sign of creative performance. Similarly, Item 4 depends on the student's judgment, and may or may not be an accomplishment which has received public recognition.

Creative Arts Scale

1. Won one or more speech contests.
2. Had poems, stories, or articles published in a *public* newspaper or magazine (not school paper) or in a state or national high school anthology.
3. Won a prize or award in an art competition (sculpture, ceramics, painting, etc.).
4. Received the highest rating in a state music contest.
5. Received one of the highest ratings in a national music contest.
6. Composed music which has been given at least one public performance.
7. Arranged music for public performance.
8. Had minor roles in plays (not high school or church-sponsored).
9. Had leads in high school or church-sponsored plays.
10. Won literary award or prize for creative writing.
11. Had cartoon published in public newspaper or magazine, etc. (not high school paper).

For boys, the estimated reliabilities (Kuder-Richardson Formula 20) of the Arts and Science scales, respectively, are .36 and .55. For girls, the reliabilities are .38 and .37. These low reliabilities may be attributed to the brevity of the scales; they limit the level of correlation between the predictors and these criteria. The Arts and Science scales correlate with one another .15 and .09 for boys and girls, respectively.

Predictors

Students and parents were sent the following inventories and scales. Students filled out:

1. The Gough (1957b) Differential Reaction Schedule (DRS), an inventory which was designed to predict originality and which contains the following scales: Intellectual Competence, Inquiringness, Cognitive Flexibility, Esthetic Sensitivity, Sense of Destiny, Total Score (sum of all scores), and Potential Success (PIV), an index of general drive and ambition derived from a variety of Institute of Personality Assessment and Research (IPAR) studies of achievement and personal effectiveness.

2. The Complexity-Simplicity, Independence of Judgment, and Originality scales from Barron's (1952) Inventory of Personal Philosophy. Only 54 of the 150 items in the Originality scale were used. The Complexity-Simplicity scale consists of items from the original 900-item IPAR inventory, selected for their significant correlations with scores on the Barron-Welsh Art Scale in five different samples, four groups of college students and one of military officers. The successive sample method was used for item selection; on cross-validation the final scale correlated .7 with the Figure Preference scale. The Originality scale was developed out of an MMPI-CPI pool by correlating each item with a composite score on four measures of originality (Guilford's Unusual Uses, Consequences, and Plot Titles; and the IPAR Word Rearrangement test) in a sample of 343 military officers. It was cross-validated by correlating it with a composite of the three Guilford tests in a sample of 150 undergraduates at the University of California.[8]

3. The Mastery scale, an 18-item scale derived from three items on Strodtbeck's Value Scale (1958).

4. The Deferred Gratification scale, formerly called the Play scale, from the National Merit Student Survey (Holland, 1960a).

5. The Vocational Preference Inventory (VPI), a short revision of the Holland Vocational Preference Inventory (Holland, 1958).

6. The Ghiselli Self-Description Inventory. Since the correlations between weighted and unweighted scale scores for the Initiative, Self-Assurance, and Occupational Level scales ranged from .89 to .96, the items of these scales were scored with unit weights rather than being weighted differentially.

7. Self-ratings on a four-point scale and a Self-Evaluation score based on the number

[8] Frank Barron, personal communication, June 1960.

of times a student rates himself above average on 20 traits.

8. The Creative Activities scale (CAS), a list of hobbies and activities assumed to demand original behavior. Teacher Ratings of Citizenship, Popularity, and Social Leadership were also available for each student, as well as SAT scores, level of degree sought, Birth Order, and a Breadth of Interest score based on the number of activities and interests checked from a list of 47 items.

Mothers filled out the PARI (Schaefer & Bell 1958), and 16 of the 23 five-item scales were used. Fathers gave information on family background and ranked nine goals and traits in the order in which they considered them valuable for their children to possess.

Results

Correlational analysis. The 75 predictors and criteria were intercorrelated (product-moment) and the complete 75 × 75 matrices for boys and girls, along with the means and standard deviations, have been deposited with the ADI.[4] Tables 1 and 2 show only the correlations between the three criteria of academic and creative performance and the 72 predictive variables. The influence of intelligence has been partialed out, using the SAT-M score as an estimate of intelligence for all three criteria, since the math score has higher correlations with the three criteria than does the verbal score. Generally, SAT-M has low correlations with the predictors. The largest absolute difference between corrected and uncorrected r's was only .05.

For boys, a review of the correlations at the 1% level reveals that the three kinds of performance (scientific, artistic, and academic) are associated with somewhat different variables, generally in expected directions.

The boy with a high score on the Creative Science Scale also has many artistic achievements (3) and Creative Activities (8), plans to get an advanced degree (71), is a first-born or an only child (74), and has high scores on the Independence of Judgment (17), Mastery (19), Deferred Gratification (20), Initiative (23), Physical Activity (28), and Intellectuality (29) scales and low scores on the Responsibility (30) and Status (36) scales. He rates himself high on Originality (37), Independence (40), and Perseverance (42). Fathers of boys high on scientific performance regard curiosity (48) as a valuable trait for their sons to have; mothers of these boys tend to be agreeable (supposedly comfortable in their role as mothers) rather than irritable (59).

The boy with a high score on the Creative Arts Scale engages in many Creative Activities (8), plans to get an advanced degree (71), and has high scores on the Esthetic Sensitivity (12), Sense of Destiny (13), DRS (Total Score) (14), Breadth of Interest (21), Self-Evaluation (22), Self-Assurance (24), Occupational Level (25), Response Bias (Acquiescence) (26), Responsibility (30), Verbal Activity (32), Emotionality (33), Control (34), and Status (36) scales, and low scores on SAT-M (5). He rates himself high on Originality (37), Drive to Achieve (39), Independence (40), Self-Confidence (41), and Perseverance (42). Mothers of these boys are slightly more accepting of the homemaking role than are mothers of stu-

[4] The following tables have been deposited with the American Documentation Institute: Tables A and B, Correlational matrices for boys and girls; Tables C and D, Means and standard deviations for boys and girls on 75 variables; and Tables E and F, Daydreams about future occupations for students who are high and low in academic and creative performance. Order Document No. 6611 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

CORRELATIONS BETWEEN THREE CRITERIA OF ACADEMIC AND CREATIVE PERFORMANCE AND 72 APTITUDE, PERSONALITY, AND BACKGROUND VARIABLES FOR BOYS
(N = 649)

Variable	Creative Performance		High School Grades	Variable	Creative Performance		High School Grades
	Scientific	Artistic			Scientific	Artistic	
1. HSR	00	08	—	39. Drive to Achieve	08*	22**	28**
2. Scientific Performance	—	15	00	40. Independence	15**	20**	06
3. Artistic Performance	15**	—	08	41. Self-Confidence	10*	23**	13**
4. SAT-V	02	−09*	−04	42. Perseverance	11**	15**	27**
5. SAT-M	04	−19**	09*				
6. Humanities Comprehension	−01	03	02	Teacher Ratings			
7. Scientific Comprehension	06	03	04	43. Citizenship	00	09*	26**
8. Creative Activities (NMSS)	36**	37**	−03	44. Popularity	01	05	17**
				45. Social Leadership	00	10*	19**
Gough DRS				Father's Values and Goals			
9. Intellectual Competence	−03	02	−01	46. Defend Self	−01	00	−02
10. Inquiringness	00	02	02	47. Ambitious	06	00	05
11. Cognitive Flexibility	02	01	−09*	48. Curious	11**	−08*	−09*
12. Esthetic Sensitivity	01	12**	01	49. Dependable	−10*	−02	02
13. Sense of Destiny	03	23**	−04	50. Good Student	−02	06	16**
14. Total Score (Above 5)	03	15**	−03	51. Happy, Well-Adjusted	−02	01	−06
15. Potential Success (PIV)	01	04	01	52. Independent, Self-Reliant	−02	04	−12**
				53. Popular	−01	01	09*
Originality-Personality				54. Self-Controlled	−01	00	−06
16. Complexity-Simplicity	07	07	−14**	PARI			
17. Independence of Judgment	10*	03	−11**	55. Fostering Dependency	−01	−06	05
18. Barron Originality	07	08*	−12**	56. Seclusion of Mother	01	−02	10*
19. Mastery	11**	10*	13**	57. Martyrdom	03	−03	09*
20. Deferred Gratification	11**	06	11**	58. Strictness	−07	03	03
21. Breadth of Interest	09*	23**	00	59. Irritability	−11**	−05	−02
22. Self-Evaluation	01	23**	24**	60. Excluding Outside Influences	03	−06	06
				61. Suppression of Aggression	07	−04	08*
Ghiselli				62. Rejection of Homemaking Role	−02	−11**	02
23. Initiative	11**	07	−02	63. Equalitarianism	−05	03	02
24. Self-Assurance	09*	12**	11**	64. Approval of Activity	00	−07	06
25. Occupational Level	01	11**	−07	65. Avoidance of Communication	05	03	12**
VPI				66. Suppression of Sex	03	−04	05
26. Response Bias	−02	13**	05	67. Ascendency of Mother	−05	−09*	05
27. Infrequency	−07	−01	03	68. Intrusiveness	10*	03	05
28. Physical Activity	18**	−04	02	69. Comradeship and Sharing	−07	−02	06
29. Intellectuality	23**	03	07	70. Acceleration of Development	07	−03	10*
30. Responsibility	−17**	20**	05				
31. Conformity	−07	01	06				
32. Verbal Activity	−10*	11**	00				
33. Emotionality	−05	31**	02				
34. Control	−02	12**	15**	Miscellaneous			
35. Masculinity-Femininity	10*	−05	03	71. Degree (level) Sought	15**	13**	04
36. Status	−18**	18**	12**	72. Father's Educational Level	05	05	−08*
				73. Mother's Educational Level	−02	03	−10*
Self-Ratings							
37. Originality	15**	27**	−01	74. Birth Order (first born)	09*	−04	01
38. Popularity	−05	10*	12**	75. School Size	−02	−04	03

* Significant at the .05 level.
** Significant at the .01 level.

TABLE 2

CORRELATIONS BETWEEN THREE CRITERIA OF ACADEMIC AND CREATIVE PERFORMANCE AND 72 APTITUDE, PERSONALITY, AND BACKGROUND VARIABLES FOR GIRLS

(N = 345)

Variable	Creative Performance		High School Grades
	Scientific	Artistic	
1. High School Grades	00	−09	—
2. Scientific Performance	—	09	−03
3. Artistic Performance	09	—	−09
4. SAT-V	07	01	08
5. SAT-M	13*	−12*	24**
6. Humanities Comprehension	03	05	07
7. Scientific Comprehension	11*	−01	05
8. Creative Activities (NMSS)	23**	36**	−19**
Gough DRS			
9. Intellectual Competence	−14**	−02	−01
10. Inquiringness	00	02	−03
11. Cognitive Flexibility	00	09	−12*
12. Esthetic Sensitivity	06	08	−01
13. Sense of Destiny	03	23**	−04
14. Total Score (Above 5)	−01	14**	−08
15. Potential Success (PIV)	−04	00	00
Originality-Personality			
16. Complexity-Simplicity	02	06	−14**
17. Independence of Judgment	08	02	−11*
18. Barron Originality	07	10	−05
19. Mastery	19**	02	08
20. Deferred Gratification	11*	03	14**
21. Breadth of Interest	18**	24**	−09
22. Self-Evaluation	08	19**	13*
Ghiselli			
23. Initiative	−03	07	03
24. Self-Assurance	03	00	15**
25. Occupational Level	01	10	01
VPI			
26. Response Bias	00	05	09
27. Infrequency	−05	−06	06
28. Physical Activity	08	−04	−03
29. Intellectuality	11*	−07	−06
30. Responsibility	02	06	06
31. Conformity	−03	−07	02
32. Verbal Activity	−06	06	11*
33. Emotionality	−02	10	07
34. Control	04	−01	14**
35. Masculinity-Femininity	06	06	−02
36. Status	−02	02	16**
Self-Ratings			
37. Originality	13*	32**	−03
38. Popularity	−01	10	04
39. Drive to Achieve	15**	13*	18**
40. Independence	07	12*	−04
41. Self-Confidence	05	11*	12*
42. Perseverance	13*	13*	22**
Teacher Ratings			
43. Citizenship	05	−04	30**
44. Popularity	−02	−04	23**
45. Social Leadership	01	04	23**
Father's Values and Goals			
46. Defend Self	08	00	00
47. Ambitious	13*	05	04
48. Curious	06	03	−16**
49. Dependable	−15**	−02	16**
50. Good Student	11*	08	−01
51. Happy, Well-Adjusted	−13*	00	−02
52. Independent, Self-Reliant	00	−09	−01
53. Popular	−11*	−04	00
54. Self-Controlled	−05	−01	02
PARI			
55. Fostering Dependency	−02	−10	04
56. Seclusion of Mother	−03	02	15**
57. Martyrdom	−06	−05	08
58. Strictness	03	−01	07
59. Irritability	−08	−07	03
60. Excluding Outside Influences	−01	−10	11*
61. Suppression of Aggression	−06	−06	16**
62. Rejection of Homemaking Role	−06	−05	06
63. Equalitarianism	−04	−06	01
64. Approval of Activity	−01	−06	12*
65. Avoidance of Communication	00	−07	13*
66. Suppression of Sex	−05	−06	10
67. Ascendency of Mother	−01	−12*	05
68. Intrusiveness	−09	−03	05
69. Comradeship and Sharing	−08	−13*	00
70. Acceleration of Development	04	−01	15**
Miscellaneous			
71. Degree (level) Sought	19**	09	−03
72. Father's Educational Level	06	10	−08
73. Mother's Educational Level	01	13*	00
74. Birth Order (first born)	−08	−02	−01
75. School Size	−03	−08	02

* Significant at the .05 level.

** Significant at the .01 level.

dents with low artistic performance (62).

The variables related to academic performance are somewhat different from those related to creative performance. Academic performance (HSR) is negatively correlated with high scores on the Complexity-Simplicity (16), Independence of Judgment (17), and Barron Originality (18) scales and positively correlated with high scores on the Mastery (19), Deferred Gratification (20), Self-Evaluation (22), Self-Assurance (24), Control (34), and Status (36) scales. Boys with high grades rate themselves high on Popularity (38), Drive to Achieve (39), Self-Confidence (41), and Perseverance (42). Their teachers rate them high on Citizenship (43), Popularity (44), and Social Leadership (45). Fathers of these students want their sons to be Good Students (50) and are less concerned with their being Independent and Self-Reliant (52). Mothers tend to be authoritarian (at the 1% level), Avoidance of Communication (65); at the 5% level, Seclusion of Mother (56), Martyrdom (57), Suppression of Aggression (61), and Acceleration of Development (70).

The findings for girls, shown in Table 2, are similar to those for boys. The girl with a high score on the Creative Science Scale tends to engage in many Creative Activities (8), plans to get an advanced degree (71), has low scores on the Intellectual Competence scale (9) and high scores on the Mastery (19), Breadth of Interest (21) scales. She rates herself high on Drive to Achieve (39). Her father ranks being Dependable (49) low on the list of goals and traits which he values for his children.

The girl with a high score on the Creative Arts Scale has high scores on the Creative Activities (8), Sense of Destiny (13), DRS (Total Score) (14), and Breadth of Interest (21) scales; has a positive Self-Evaluation (22), and rates herself high on Originality (37).

Academic performance is negatively correlated with high scores on the Creative Activities (8) and Complexity-Simplicity (16) scales and positively correlated with high scores on SAT-M (5) and the Deferred Gratification (20), Control (34), Status (36), and Self-Assurance (24) scales; with high Self-Ratings of Drive to Achieve (39) and Perseverance (42), and high Teacher Ratings of Citizenship (43), Popularity (44), and Social Leadership (45). Fathers of girls who get high grades value being Dependable (49) and do not value being Curious (48). Mothers of these girls are characterized by the PARI as "authoritarian"—Seclusion of Mother (56), Supression of Aggression (61), Acceleration of Development (70).

The differences between creative and academic performance for the 72 variables replicate the trends found for boys. Academic performance appears to be the function of a personal syndrome characterized by perseverance, self-control, good behavior (good citizenship), and rigidity; this personality pattern is related to parental attitudes which seem conducive to such a pattern. Creative performance, on the other hand, seems to be the outcome of a conscious conception of being original, active participation in creative hobbies, and reinforcement by parents who possess values and attitudes which appear to be conducive to such performance.

Free response analysis. As a part of the assessment, students were asked about their vocational choice, the factors which influenced this choice, and their images and daydreams about their future occupations. Their free

responses to the question, "What do you daydream about when you think of your future occupation?" were coded, using 10 coding categories developed from a sample of 200 students and tested for reliability on a sample of 160 students. Two raters, coding independently, obtained 75% agreement. Tables E and F[4] compare the responses of students high on creative performance with those of students low on creative performance, and of students high on academic performance with students low on academic performance.

According to Table E, boys with outstanding scientific performance daydream of "high achievement and creative accomplishment"; for this scoring category, the percentage difference between high and low science scores is significant at the .01 level. The same trend is observable in the Arts scale, though the percentage difference is not significant at the .05 level. The responses of students above the median on HSR do not differ significantly from those of students below the median.

Girls with outstanding scientific achievement daydream more about work activities than about helping others ($p < .001$), whereas girls with low achievement are less concerned with work and more interested in helping others ($p < .05$). Again, the comparisons of girls high and low on academic achievement do not reveal significant differences.

These results, which support some of the correlational findings, imply that academic achievement involves somewhat different motives than creative performance; good grades in high school appear to be a function of socialization (citizenship and popularity) and perseverance, whereas creative performance is a function of conscious concern with high accomplishment, independence, and originality.

Getzel's analysis. In a recent study of elementary school children, Torrance (1959), following Getzel's classification of students as "high IQ" or "high creative," reports that the most intelligent students (top 20%) are not necessarily the most creative (top 20%): only about 30% of the students were in both groups.

The present study supports Torrance's findings. Eight tables were obtained for the present study by relating SAT-V and SAT-M to the scientific and artistic checklist scores for each sex. Within each table, distributions of creative performance scores were formed for approximately the top 20% of students on the SAT and the top 20% of students on creative performance. The percentages of overlap between these two contrasting student groups ranged from 1–6.5% of the total samples for the eight tables and deviate only slightly from an expected overlap of 4%. The larger deviations were in the direction expected from the correlations in Tables 1 and 2. These results are consistent with the negligible correlations between aptitude and outstanding creative achievement shown in Tables 1 and 2. We may conclude, then, that in a group of exceptionally bright students, intelligence has little or no relationship to creative performance in arts and science or to academic achievement.

These results may in part be a function of the regression of creativity on aptitude; students are selected initially for their high aptitude, and then the relationships between aptitude and creativity are studied within this narrow range of aptitude. These low relationships are, only in part, a function of the narrow range on aptitude, however. In Tables A and B of the ADI material SAT-V and SAT-M

correlate as much as .64 and .57 with the Humanities and Scientific Comprehension scores of the National Merit Scholastic Scholarship Test (an achievement test given about one year before the SAT was administered). Even these correlations are probably underestimates, due to the artificial restriction of range imposed on the NMSQT by its use as a selection device for this sample. The latter findings suggest that the SAT does discriminate efficiently and reliably within this narrow range of talent so that higher relationships between aptitude and creative as well as academic performance were possible.

Discussion

The results of the present study are probably conservative. The distribution of scores for all criteria are at least moderately skewed and all scales have relatively low reliability; moreover, the range of scores on the remaining predictor variables, including background variables, is restricted. Therefore, the correlational results are probably underestimated. Since the students in the sample are all exceptionally talented, the correlations between originality measures and scholastic aptitude are reduced. Whether or not the observed relationships can be generalized to more representative student populations is questionable.

The correspondences between the present study and related studies are numerous. The originality scales developed by Barron (Independence of Judgment, Complexity-Simplicity, and Originality) and by Gough (DRS) have been frequently correlated with criteria of creative performance and our results are generally consistent with earlier reports (Barron, 1953a, 1953b; Gough, 1957b). Although the relationships obtained in this study are generally low, it is surprising that the Barron and Gough scales correlate at all, since they were constructed for adult samples and used different criteria of creative behavior.

The CAS has promising validity since it is the best single correlate of creative performance (*r*'s range from .23–.37 for the CAS against the Arts and Science criteria for both sexes), and since its highest correlations are with a set of variables which lend support to this interpretation. For boys, the variables having the largest correlations with the CAS are, in descending order (.37–.15): Breadth of Interest, Self-Ratings of Originality and Independence, Emotionality, Self-Ratings of self-control, Total DRS Score, Complexity-Simplicity, level of degree sought, and Barron Originality. For girls, the largest correlates of the CAS are (.43–−.16): Breadth of Interest, Self-Ratings of Originality and Independence, Barron Originality, Complexity-Simplicity, Sense of Destiny, Self-Evaluation, and mother's education. The Fostering Dependency and Ascendency of Mother scales of the PARI correlate negatively with the CAS (−.16 and −.17, respectively).

The assumed validity of the CAS may be overestimated, since 4 or 5 of the 32 scale items involve activities which are preparatory to or associated with the accomplishments listed in the Creative Science and Arts Scales. The correlates of the CAS listed above imply some validity independent of the CAS's correlations with the performance scales, however.

The CAS was suggested by the Chorness and Nottelmann study (1957) in which it was found that a set of extracurricular hobbies had significant correlations (.27–.45) with four or five Guilford factor composites when intelligence is partialed out. The item overlap, if any, between the

Chorness creative hobbies and the CAS is not known.

The unexpected concurrent validity of the Self-Ratings is of special interest; of all the variables used, the Self-Ratings show most clearly that academic and creative performers conceive of themselves as different from other people. It is of interest too that high Self-Ratings of Perseverance and Self-Confidence—two of the three most discriminating traits for Terman's high and low achievers—are correlated here with both creative and academic performance (Terman & Oden, 1947).

The findings for the Intellectuality and Emotionality scales of the VPI appear to replicate several studies (Barron, unpublished; MacKinnon, 1959) in which vocational interests in science and the arts were found to be associated with rated creative performance and high scores on tests of creative behavior. Since the Intellectuality scale has its highest significant correlations with the Self-Sufficiency, Schizothymia, and Masculinity scales of the 16 PF, the present results are consistent with the characterization of the creative person as independent, asocial, and masculine. Similarly, the Emotionality scale of the VPI is correlated most highly with femininity, instability, and introversion (16 PF) Holland, 1960b).

Generally, the results seem to support many other hypotheses about the nature of the creative person, but an attempt to relate all of the specific findings to the voluminous theorizing about creativity would be a gargantuan task. Briefly, those hypotheses which regard the creative person as independent, intellectual, expressive, asocial, consciously original, and open to experience gain some support (Taylor, 1955, 1957, 1959). The creative student not only has an identifiable personal disposition, but also he has relationships with teachers and parents which are in accordance with our expectations. His parents appear to be more permissive and more nurturant of his ideas and impulses so that communication with the self and the world is stimulated. It is not surprising, then, that such students appear to come in conflict with teachers who demand conforming, controlled, nonexploratory behavior.

Perhaps the most unequivocal finding in the present study is that, for samples of students of superior scholastic aptitude, creative performance is generally unrelated to scholastic achievement and scholastic aptitude. Since the traditional predictors of scholastic aptitude are of little or no value for predicting creativity, it seems clear that scholarship programs, colleges, and other agencies, if they are concerned primarily with rewarding students or selecting employees who have potential for creative performance, need to make a more active effort to devise predictors of creative potential. In fact, attempts to build better scholastic aptitude tests may even be detrimental, since they may lead to a greater dependence on instruments which are of limited value and thus delay unnecessarily the development of efficient predictors of creative performance.

The finding that teachers generally rate students with good grades higher than those with lower grades, together with the results of an earlier study (Holland, 1959b) which indicates that teachers give lower ratings to students with more potential for creativity (as measured by the 16 PF) than to students with less potential, suggests that school personnel value the good grade-getter more than the creative student. Recently, Jex (1958) found that school principals rate teachers with high

scores on an ingenuity test lower than teachers with low test scores. Taken together, these results reinforce a common belief that some of our traditional education may be stultifying rather than nurturant and fructifying.

Summary

The relationships between three criteria of academic and creative performance and 72 personal, demographic, and parental variables were studied in a sample of talented adolescents. The results suggest that creative performance at the high school level occurs more frequently among students who are independent, intellectual, expressive, asocial, consciously original, and who have high aspirations for future achievement. Students who are persevering, sociable, responsible, and whose parents hold somewhat authoritarian attitudes and values, are more frequently academic achievers. The negligible relationships found between academic aptitude and creative performance at a high aptitude level suggest that we need to use nonintellectual criteria in the selection of students for scholarships and fellowships.

REFERENCES

Barron, F. *Inventory of personal philosophy.* Berkeley, Calif.: Univer. California Press, 1952. (Institute of Personality Assessment and Research)

Barron, F. Complexity-simplicity as a personality dimension. *J. abnorm. soc. Psychol.*, 1953, **48**, 163–172. (a)

Barron, F. Some correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287–297. (b)

Chorness, M. H., & Nottelmann, D. N. The prediction of creativity among Air Force civilian employees. *USAF Personnel Train. Res. Cent. res. Rep.*, 1957, No. AFPTRC-TN57-36.

d'Heurle, Adma, Mellinger, Jeanne C., & Haggard, E. A. Personality, intellectual, and achievement patterns in gifted children. *Psychol. Monogr.*, 1959, **73** (13, Whole No. 483).

Drews, Elizabeth M., & Teahan, J. E. Parental attitudes and academic achievement. *J. clin. Psychol.*, 1957, **13**, 328–332.

Gough, H. G. What determines the academic achievement of high school students? *J. educ. Res.*, 1953, **46**, 321–331.

Gough, H. G. Imagination: Undeveloped resource. Paper read at Proceedings of the first conference on research and development in personnel management, Los Angeles, California, 1957. (a)

Gough, H. G. Researchers' summary for the Differential Reaction Schedule. Univer. Calif., 1957. (Ditto) (b)

Holland, J. L. A personality inventory employing occupational titles. *J. appl. Psychol.*, 1958, **42**, 336–342.

Holland, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *J. educ. Psychol.*, 1959, **50**, 135–142. (a)

Holland, J. L. Some limitations of teacher ratings as predictors of creativity. *J. educ. Psychol.*, 1959, **50**, 219–223. (b)

Holland, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245–254. (a)

Holland, J. L. The relation of the Vocational Preference Inventory to the Sixteen Personality Factor Questionnaire. *J. appl. Psychol.*, 1960, **44**, 291–296. (b)

Jex, F. B., & Merrill, R. M. An evaluation of the first academic-year institute, University of Utah, [illegible]. *U. Utah res. Monogr. Educ.*, [illegible], No. 2.

MacKinnon, D. W. What do we mean by talent and how do we test for it? In, *The search for talent: College admissions 7.* New York: College Entrance Examination Board, 1959.

Schaefer, E. S., & Bell, R. Q. Development of a Parental Attitude Research Instrument. *Child Develpm.*, 1958, **29**, 339–361.

Strodtbeck, F. L. Family interaction, values, and achievement. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and society.* New York: Van Nostrand, 1958. Ch. 4.

Taylor, C. W. (Ed.) *The first, second, and third University of Utah research con-*

ferences on the identification of creative scientific talent. Salt Lake City: Univer. Utah Press, 1955, 1957, 1959.

TERMAN, L. M., & ODEN, MELITA H. *Genetic studies of genius: IV. The gifted child grows up*. Stanford: Stanford Univer. Press, 1947.

TORRANCE, E. P. Explorations in creative thinking in the early school years: A progress report. In, *The third University of Utah research conference on the identification of creative scientific talent*. Salt Lake City: Univer. Utah Press, 1959. Pp. 58–71.

(Received September 26, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 148–152

BEHAVIORS DURING PROBLEM SOLVING AMONG CHILDREN OF LOW, AVERAGE, AND HIGH INTELLIGENCE[1]

HERBERT J. KLAUSMEIER AND LEO J. LOUGHLIN
University of Wisconsin *Northern Illinois University*

The range of differences among school children in many intellectual and psychomotor abilities and other characteristics is defined clearly. Less clearly defined are the factors which contribute to the differences. Previously, no statistically significant difference was found among groups of low, average, and high IQ children in retention of arithmetic problems, graded to each child's achievement level (Klausmeier & Feldhusen, 1959); nor in emotional adjustment, integration of self-concept, and achievement in relation to expectancy (Klausmeier, Check, & Feldhusen, 1960). Large differences with little overlapping between IQ levels were found in achievements in reading, language, and arithmetic. The purpose of this investigation was to ascertain possible differences among the three groups in efficiency of method, persistence, and mode of attack while in the actual process of solving arithmetic problems.

PROCEDURE

Subjects

The subjects were 20 boys and 20 girls with WISC IQs of 55–80, 20 boys and 20 girls with WISC IQs of 90–110, and 20 boys and 20 girls with WISC IQs of 120–146. The 80 children of high and average IQ were enrolled in 15 regular fifth grade classes of two large public school systems; and the 40 low IQ children were enrolled in 14 special classes for educable mentally retarded children in the same schools. Excluded from the study were low IQ children exhibiting a second handicap to a marked degree or definite organic symptoms of retardation such as mongolism. All had birthdates between September 15, 1947, and December 15, 1948; and the mean age at the time of gathering the data was 131 months for each IQ group.

Methods

The first task was to get problems properly graded in difficulty. After extensive experimenting with pupils in regular classrooms and with low IQ pupils not in the study, a group of 29 problems at three main levels of difficulty was arranged. The problems for the low IQ children were subdivided further with the lowest level intended for those who did not know the number of cents in a dime or who could not perform simple addition. A reserve level of problems of high difficulty was established for the high IQ child who might solve an assigned problem in less than 3 minutes The least difficult problem designed for a low IQ child was to make 7¢ with three coins and the most difficult designed for a high IQ child was to make $14.53 with 16 bills and coins.

Besides this initial formulation of problems, two other measures were used to make certain that the problem was of appropriate difficulty. A survey test was administered to each child to ascertain his knowledge of our money system, related vocabulary, and mathematical skills. Also, if a child completed his assigned problem in less than 3 minutes, the next problem in order of difficulty was assigned.

Each child solved only one problem. This problem and the survey test mentioned previously were administered by a carefully trained examiner; and another trained person observed the entire procedure. From the time the problem and instructions were given the child until the problem was solved, time in seconds and the incidence of the following behaviors of the child were recorded: (*a*) manipulations of coins by the low IQ children or manipulation of figures

[1] The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare.

with pencil and paper by the average and high IQ children; (*b*) engaging in nontask activities or withdrawal which resulted in the examiner's encouraging the child to continue his problem solving; (*c*) offering an incorrect solution; (*d*) noting and correcting mistakes independently; (*e*) verifying the solution by checking the amount of money independently; (*f*) verifying the solution by checking the number of coins independently; and (*g*) using a random, logical, or a combination approach to the solution. In random approach the subject merely manipulated or explored the materials without an apparent plan of attack. In logical approach the subject made deliberate manipulations to test successive trials or hypotheses. Some children changed from an apparently random to a logical approach, and vice versa.

Efficiency of the child's method was defined as the ratio of moves (coins or paper-pencil computations made) to the least number needed to achieve the solution. For example, the least efficient subject made 39 moves when only 2 were required, with a resulting efficiency ratio of .051. Any instance of a child's being encouraged by the examiner to cease from a nontask activity or to continue at the task was recorded as the measure of nonpersistence. The mode of attack was observed in the verbal responses of the subject and in his manipulations of coins or figures, as described in the preceding paragraph, *c* through *f*.

The data secured on each subject were tabulated and combined according to the IQ groups. Analysis of variance was used to compare efficiency of method and chi square to compare incidences of the other behaviors among the three groups.

Results

The extent to which the problems were of equal difficulty for each IQ group was ascertained by comparing the time in seconds required to reach a solution to the problem. The low IQ group required a mean time of 624.17 seconds, the average group 627.65, and the high group 578.47 seconds. On the basis of a small and insignificant *F* value obtained among the means, it was concluded that the assigned problems were about equally difficult for the three IQ groups.

A summary of the data and the results of the tests of statistical significance of differences among the three IQ groups and between groups are presented in Table 1. Column 1 lists the observed behaviors. Columns 2–4 contain the scores or number of subjects in each IQ group who demonstrated the behavior. Column 5 shows the results of the tests of statistical significance of differences among the three IQ groups. Columns 6–8 give the results of the tests of sig-

TABLE 1
SUMMARY OF OBSERVED SCORES AND RESULTS OF TESTS OF STATISTICAL SIGNIFICANCE OF DIFFERENCES

Behavior	Score or Incidence			Significance of Differences			
	High	Average	Low	Among Groups	Between Groups		
					Hi-Av	Hi-Low	Av-Low
Efficiency of Method (%)	.353	.309	.237	$.05 > p > .01$	ns*	$.05 > p > .01$	ns
Nonpersistence (Incidence)	3	10	17	$.01 > p > .001$	$.02 > p > .01$	$.01 > p > .001$	$.02 > p > .01$
Mode of Attack (Incidence)							
Offered Incorrect Solution	10	16	33	$.001 > p$	$.001 > p$	$.001 > p$	$.01 > p > .001$
Noted and Corrected Mistakes	14	12	5	ns	—	—	—
Verified Solution—Amt. Money	38	31	11	$.001 > p$	$.001 > p$	$.001 > p$	$.001 > p$
Verified Solution—No. Coins	36	26	9	$.001 > p$	$.001 > p$	$.001 > p$	$.001 > p$
Random Approach	1	8	21	$.001 > p$	$.01 > p > .001$	$.001 > p$	$.001 > p$
Logical Approach	14	5	3	$.01 > p > .001$	$.01 > p > .001$	$.01 > p > .001$	$.05 > p > .02$

* Not significant at .05 level.

nificance of differences between groups —high and average, high and low, and average and low.

Table 1 shows that the low IQ group was inferior to the high IQ and to the average IQ group in seven and six of the observed behaviors, respectively; the average IQ group was inferior to the high IQ in six behaviors. Although the mean number of seconds to solve was nearly identical for the three groups, the incidences of behaviors were greatly different. To clarify the differences, short sketches of the procedures used by a high and a low IQ boy are now given.

Richard had a WISC IQ of 138 and used 585 seconds to solve the problem: "Make $9.77, using 12 bills and coins." He studied the situation and during the first 3 minutes made a first attempt at solution, getting the amount correct but using 13 instead of 12 bills and coins—9 one dollar bills, 1 half-dollar, 1 quarter, and 2 pennies. He checked the amount and number of coins and started over. During the next 3 minutes he made his second and third tries, each time starting with a five dollar bill. On the second he had 12 bills and coins but an incorrect amount; on the third try, neither was correct. Each time he checked amount and number of coins. During the third 3 minutes he made his fourth try, but made a mistake in adding the amount. Checking through his solution at the beginning of the fourth 3-minute period, he found the error and after 45 seconds offered as a correct solution 1 five dollar bill, 4 one dollar bills, 1 half dollar, 1 dime, 3 nickels, and 2 pennies. Richard made 25 separate moves, but only 6 were needed, as shown in the final solution. Therefore, his efficiency ratio was 6/25, or .24.

In terms of a problem solving sequence set forth by Merrifield, Guilford, Christensen, and Frick (1960), Richard (*a*) accepted the task as a problem and set a goal of achieving a correct solution, (*b*) analyzed the available information in relation to his goal, (*c*) produced a solution and independently rejected it (verification), (*d*) applied knowledge from this and subsequent solutions to each successive solution produced (reapplication), and (*e*) finally verified and accepted the correct one.

Michael, one of the more able low IQ boys, had a WISC IQ of 73 and used 805 seconds to "Make 46¢, using 4 coins." Ten coins of each demonination—penny, nickel, dime, and quarter—had been placed on the table before him. (In a pretest it had been established that he had the requisite skill in adding and knowledge of coins to solve the problem.) Michael, as Richard, restated the problem correctly prior to attempting a solution. During the first 3 minutes, Michael took 3 dimes and 10 pennies from the supply, then returned to the supply 3 pennies, one at a time, and offered the remainder as a correct solution. Upon being told that it was incorrect and with all coins returned to the supply, he studied the coins and offered 1 quarter and 2 dimes as a second solution during the second 3 minute period. During the third 3 minutes, he withdrew from the supply 1 quarter, 1 dime, 1 nickel, and 1 penny and offered this as correct. When told it was incorrect and the coins had been returned, he again studied the coins, and during the fourth 3 minutes withdrew 1 quarter, 1 penny, 1 dime, then put the quarter back. During the fifth 3 minutes he again withdrew the quarter and another dime from the supply, thus having 1 quarter, 2 dimes, and 1 penny. This he offered as correct without observably checking ac-

curacy of amount or number of coins. He made a total of 16 coin movements when only 3 were needed; his efficiency ratio was .188. At no point was it necessary to encourage Michael to continue or to keep at the task; however, like the majority of low IQ children, he offered incorrect solutions and did not verify amount of money nor the number of coins.

In what respects are the low IQ most different from the high IQ children? As noted in the above descriptions, the high IQ boy was markedly superior in verification and reapplication, eventually arriving at the correct solution. Further, other low IQ children, different from Michael, did not persist when informed that their solutions were wrong, and after two or three such experiences, manipulated the coins without producing a close approximation of the final correct solution.

Discussion

If time to reach a correct solution had been used as the criterion of efficiency of the problem solving, no significant difference would have been observed among the three groups. Using the ratio of efficiency as the criterion, a significant difference was observed. The differences of most practical significance among the groups were connected with producing possible solutions, verifying each proposed solution, and applying information from each preceding trial to the subsequent one. As John (1957), Duncan (1959), and others have pointed out, processes such as these used in arriving at solutions are the important variables in understanding efficiency of performance and learning.

Whether more of the low IQ children in this study could learn seemingly simple processes as deciding a plan for making 13¢ with four coins is not known. Similarly, verifying the amount of money and number of coins does not appear to be difficult, although the high IQ children also tended to do only one or the other on the first try. If these processes can be learned, more specific instruction in school devoted to helping children acquire processes, rather than solutions, would be helpful.

Though not studied intensively, differences within each IQ group were large. For example, the ratio of efficiency varied from .10 to .80 among the high IQ children. Further, 10 high and 16 average IQ children offered incorrect solutions as shown in Table 1. Help given to children of average and high abilities when needed might be especially profitable in achieving improved problem solving performances.

Summary

Differences among groups of children of low, average, and high IQ were found in observed behaviors as they solved problems, graded in difficulty to their present achievement levels. The high IQ children showed a greater incidence than the average and the low, and the average a greater incidence than the low in noting and correcting mistakes independently, verifying solutions, and using a logical approach; whereas the low IQ showed a greater incidence than the average and the high, and the average a greater incidence than the high in nonpersistence, offering an incorrect solution, and using a random approach. The high IQ children were superior to the low in efficiency of method, as ascertained by the ratio of number of moves made to the least number required for solution of the problem. Differences in performances

among individuals within IQ groups were also large.

REFERENCES

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

JOHN, E. R. Contributions to the study of the problem-solving process. *Psychol. Monogr.*, 1957, **71**(18, Whole No. 447).

KLAUSMEIER, H. J., CHECK, J., & FELDHUSEN, J. Relationships among physical, mental, achievement, and personality measures in children of low, average, and high intelligence at 125 months of age. *Amer. J. ment. Defic.*, 1960, **65**, 69–78.

KLAUSMEIER, H. J., & FELDHUSEN, J. Retention in arithmetic among children of low, average, and high intelligence at 117 months of age. *J. educ. Psychol.*, 1959, **50**, 88–92.

MERRIFIELD, P. R., GUILFORD, J. P., CHRISTENSEN, P. R., & FRICK, J. W. A factor-analytic study of problem-solving abilities. *Rep. Psychol. Lab., U. Southern Calif.*, 1960, No. 22.

(Received October 3, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 153–156

THE RELATION BETWEEN SEVERAL PARENT MEASURES AND THE CHILD'S EARLY ADJUSTMENT TO SCHOOL[1]

GENE R. MEDINNUS

San Jose State College

The present paper reports data dealing with first grade readiness and adjustment in first grade. The purpose of this study was to explore the relationship between parental attitudes toward a number of aspects of child rearing, as assessed through questionnaires administered to both parents and through interviews with the mother, and the child's general adjustment in Grade 1.

METHOD

Subjects

The subjects in the investigation were a group of first graders and their parents. Two groups of children were identified through the teacher's rating of their adjustment in first grade on a five-point scale, with 5 representing good adjustment, 3 typical of the adjustment of the "average" child, and 1 denoting poor adjustment. The rating was made on the basis of and following the teacher's rating of the children on a 52-item First Grade Adjustment Scale. Those 10 children receiving ratings of 4 and 5 were treated as one group; the other group was comprised of those 9 with ratings of 1 and 2. The mean IQ of the children was 111.6 according to the Stanford-Binet (Form L) intelligence tests administered by the investigator. The socioeconomic status of the families was approximately evenly divided between the upper-lower and lower-middle classes according to Warner's Index of Status Characteristics (Warner, Meeker, & Eells, 1949). At the beginning of the study the mean ages of the mothers and fathers were 32.1 and 34.1 years, respectively, and the corresponding mean years of education were 12.0 and 12.2.

[1] This research was completed under the sponsorship of the Elizabeth McCormick Memorial Fund of Chicago.

Scales

The two instruments employed to assess parental attitudes and behavior will be described first. Data from these two measures were obtained during the year prior to the child's entrance into first grade.

The parents were asked to complete a form of the Parent Attitude Research Instrument (PARI) developed at the National Institute for Mental Health. The development of the scale, its uses and reliability coefficients have been described in some detail by Schaefer and Bell (1955, 1958). Briefly, the scale is a Likert-type questionnaire consisting of a number of subscales each concerned with various aspects of family life, child rearing attitudes, and relationships. The particular forms of the scale employed consisted of 29 subscales for the mothers and 25 for the fathers. The responses to the items were scored either 4, 3, 2, or 1, depending upon whether the response was "Strongly agree," "Somewhat agree," "Somewhat disagree," or "Strongly disagree," respectively. The questionnaires were left with the parents following an evening home visit by the investigator. They were returned as they were completed.

The Fels Parent Behavior Rating Scales originally were designed in an attempt to identify areas of the home environment which might be expected to bear some causal relation to the development of the child's personality (Baldwin, Kalhorn, & Breese, 1949; Champney, 1941). Lorr and Jenkins (1953) further analyzed the intercorrelations of the Fels scales presented by Roff (1949) in an attempt to determine second-order factors of a broad character which would be useful for conceptualizing the parent-child relationship. Three factors emerged; for convenience in the present discussion these will be labeled as follows: (*a*) Dependence vs. Independence Encouraging, (*b*) Democratic vs. Authoritarian, (*c*) Degree of organization in the home. Those five scales having the highest loadings on each of these three factors were employed in the present study. The ratings were made

by the investigator on the basis of home visits and interviews with the mothers. Interrater reliability was established through previous ratings of 10 homes by the investigator and a graduate assistant.[2] The mean reliability of these ratings for the 15 scales was .80. A list of the 15 rating scales is provided in Table 1.

For the purpose of assessing the subjects' adjustment to the first grade situation, a First Grade Adjustment Scale was developed (Medinnus, in press). Briefly, the following procedure was employed in the construction of the scale. Twenty-five first grade teachers were interviewed concerning factors related to a child's adjustment in Grade 1. The teachers were asked to mention specific traits, abilities, and items of behavior which, in their opinions, characterize or describe good adjustment and poor adjustment in the first grade. The characteristics listed for good and poor adjustment were examined for parallel items which seemed to define opposite ends of a single continuum. Those items for which there was at least a minimum level of consensus were placed on a five-point scale, with 3 representing the midpoint or rating of the "average" child, and 1 and 5 describing poor and good adjustment, respectively. The 52 items were grouped into five major areas: Physical Status, Social Behavior, Emotional Behavior, Intellectual Abilities and Behavior, and Adjustment to Classroom Membership and Requirements. Interrater reliability coefficients for the five sections and for the full scale clustered about .75.

Results

Those PARI scales which differentiated significantly between the parents of the two groups of children will be mentioned first. The fathers and mothers were compared separately since the subscales and the items were not identical for the two forms. The following scales differentiated between the mothers of the well- and poorly adjusted first graders at the .05 level or beyond: Approval of Activity, Irritability, Dependency of Mother, and Fear of Harming Baby. Mothers of the well adjusted children showed more of the attributes measured by the first three scales than did the mothers of the poorly adjusted children; the reverse was true for the fourth scale mentioned. No scales reached this significance level for the fathers.

The results for the Fels scales are given in Table 1.

Discussion

If there is any validity to the premise that the child's behavior and personality are in part causally determined by parental attitudes one would expect wide differences between the attitudes of parents of well- versus poorly adjusted first grade children. That such was not revealed in the present analysis is apparent from the fact that only four scales differentiated between the two groups of mothers and none of the scales reached the significance level for the fathers. One possible interpretation of this finding emphasizes the lack of validity in the parent attitude scale employed. Another interpretation assumes that there is actually no relation between parental attitude and the child's general adjustment in the school situation. The findings with regard to the Fels scales seem to furnish evidence to discount this interpretation.

It is apparent from Table 1 that while only 5 of the 15 Fels scales differentiated between the home environments of the two groups of children, 3 of the 5 were included in the Dependence vs. Independence Encouraging factor. In addition, the parents of the well adjusted group were rated higher on the Solicitousness variable but the difference was not statistically significant. It should be noted also that in each case the homes of the well adjusted children were rated higher in the dependence-encouraging direction. While this would seem to run counter

[2] The writer is grateful to Floyd Ayers for assistance with ratings.

TABLE 1

RATINGS OF HOMES OF TWO ADJUSTMENT GROUPS

Fels scales	Well Adjusted		Poorly Adjusted		p
	M[a]	σ	M	σ	
A. Dependence vs. Independence Encouraging					
4.1 General babying	51.3	7.91	42.8	7.81	.05
1.91 Child centeredness of home	55.3	7.85	42.1	6.28	.01
7.1 Solicitousness for welfare	51.0	7.82	45.1	11.69	*ns*
4.2 General protectiveness	52.7	6.42	40.9	8.48	.01
2.12 Intensity of contact with mother	50.6	8.93	49.0	10.99	*ns*
B. Autocratic vs. Democratic					
3.15 Nondemocracy of policy	46.4	8.14	52.6	12.87	*ns*
3.22 Coerciveness of suggestion	47.6	6.92	48.7	12.42	*ns*
3.14 Nonjustification of policy	44.8	6.22	56.3	10.07	.01
3.11 Restrictiveness of regulations	47.9	8.31	47.3	18.91	*ns*
8.2 Emotionality toward child	47.5	8.06	53.2	12.93	*ns*
C. Organization and Effectiveness of Control					
3.12 Readiness of enforcement	48.8	8.15	50.1	13.37	*ns*
3.13 Severity of actual penalties	47.2	6.57	52.8	12.97	*ns*
3.3 Accelerational attempt	55.7	7.86	40.7	5.48	.001
1.7 Coordination of household	49.0	7.90	45.4	10.73	*ns*
1.2 Activeness of home	47.6	8.18	54.9	11.78	*ns*

[a] The raw score ratings were converted to McCall *T*s.

to the notion of the need for independent behavior in the school setting, the findings obtained here concerning the three scales, General babying, Child centeredness of the home, and Protectiveness, are in accord with the results of other investigations (Baldwin et al., 1949; Bandura & Walters, 1959). In their interpretation of the parent behavior ratings, the Fels researchers identified an "indulgence" pattern consisting of three variables, Babying, Protectiveness, and Solicitousness. High ratings on these variables were interpreted as indicating either one of two processes in the home: "(a) the mother is warmly doting and protective of the child; or (b) she is anxious and restrictive lest the child endanger himself or discommode her" (Baldwin et al., 1949, p. 8). On the basis of the relationship of these three variables to other rating scales, these investigators concluded that "there is a definite tendency for those parents who are rated high on indulgence to be warm, emotional, and unable or unwilling to give the child either emotional or physical freedom" (p. 8).

While two possibilities were described to explain high ratings on the indulgence variables, the burden of the present discussion is to argue that, in the present study at least, the higher ratings of the homes of the well adjusted first graders on these variables indicate acceptance of the child while the lower ratings of the homes of the poorly adjusted children point to rejection of the child. Great concern with encouraging independence may actually reflect parental rejection of the child. Parental declarations that

the child must learn to "stand on his own two feet" and "fight his own battles" may be attempts to justify nonnurturant and nonsupporting behavior growing out of feelings of rejection.

This interpretation is similar to that of Bandura and Walters (1959) who found that the parents of aggressive boys tended to discourage dependency behavior in their sons and that parental rejection was related to nonnurturant parental behavior. The authors state: "Indeed, parental rejection seems to involve a pervasive form of punishment of the child's attempts to gain emotional support from his parents" (p. 68).

Summary

The present paper presented data concerning the relationship between various parental attitudes and behavior and the child's adjustment in first grade.

A comparison was made of the parent attitude scores and Fels home ratings of two groups of children, 10 well adjusted first graders and 9 poorly adjusted first graders.

The following results emerged from the study:

1. Four of the 29 PARI subscales differentiated significantly between the mothers of the groups. None of the scales reached the .05 level for the fathers.

2. Five of the 15 Fels scales differentiated between the home environments of the two groups of children. Three of the 5 scales were included in the Dependence vs. Independence Encouraging factor.

The results were interpreted as indicating that the lower ratings of the homes of the poorly adjusted children on the several dependence-encouraging variables reflected parental rejection which was a causal factor in the children's poor adjustment to the demands of the first grade situation.

REFERENCES

Baldwin, A. L., Kalhorn, Joan, & Breese, Fay H. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, **63**(4, Whole No. 229).

Bandura, A., & Walters, R. H. *Adolescent aggression*. New York: Ronald, 1959.

Champney, H. The measurement of parent behavior. *Child Develpm.*, 1941, **12**, 131–166.

Lorr, M., & Jenkins, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, **17**, 306–308.

Medinnus, G. R. The development of a first grade adjustment scale. *J. exp. Educ.*, in press.

Roff, M. A factorial study of the Fels Parent Behavior Scales. *Child Develpm.*, 1949, **20**, 29–45.

Schaefer, E. S., & Bell, R. Q. Parental Attitude Research Instrument (PARI): Normative data. Unpublished manuscript, National Institutes of Health, 1955.

Schaefer, E. S., & Bell, R. Q. Development of a Parental Attitude Research Instrument. *Child Develpm.*, 1958, **29**, 339–361.

Warner, W. L., Meeker, Marchia, & Eells, K. *Social class in America*. Chicago: Science Research Associates, 1949.

(Received October 3, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 157-160

ON THE USE OF A SHORT VOCABULARY TEST TO MEASURE GENERAL INTELLIGENCE

JOHN B. MINER
University of Oregon

In a prior publication (Miner, 1957) the author summarized the results of some 21 studies which yielded estimates of the relationship between vocabulary tests and measures of what we have come to call *general intelligence*. Most of these studies utilized samples characterized by a rather wide range of ability and employed vocabulary measures of intermediate length such as those contained in the Wechsler or Stanford-Binet tests. The general measures were the 1916 Stanford-Binet, the 1937 Stanford-Binet, the Wechsler-Bellevue scales, the Wechsler Intelligence Scale for Children, the General Classification Tests developed by the Armed Services, and the Otis tests. The median correlation found was .83. On the basis of this evidence, plus the knowledge that the various general measures themselves do not intercorrelate at a much higher level, it was assumed that general intelligence can be adequately predicted from a vocabulary test.

The studies of intelligence differences between various population subgroups and of manpower utilization within the educational and occupational systems of this country which are presented in *Intelligence in the United States* (Miner, 1957) are based on findings obtained with a short vocabulary test. The particular test used was found in three separate investigations (Hagen & Thorndike, 1955; Thorndike, 1942; Thorndike & Gallup, 1944) to have a reliability in the .80 to .85 range. In view of this evidence of reliability and existing knowledge as regards the relationship between vocabulary and general intelligence, the assumption that an adequate measure of group differences in intelligence was being employed seemed justified.

This conclusion has, however, been criticized on several grounds, perhaps with some justification. The short vocabulary test contained only 20 items whereas the Wechsler and Stanford-Binet measures are considerably longer. Furthermore, the short vocabulary test used a multiple-choice format rather than a free response procedure. A question was raised as to whether the specific vocabulary test used in the studies presented in *Intelligence in the United States* was in fact as valid a measure of intelligence as the presumptive evidence would suggest. No information on the correlation between the short vocabulary test and a more general measure was presented in the book, nor was such information available to the author at the time the book was published.

The present work was carried out with a view to remedying this deficiency and casting further light on the validity of the conclusions reached in *Intelligence in the United States*.

Procedure and Results

In connection with a study designed to investigate sales effectiveness, a number of tests were administered to 108 members of the sales organization of a large corporation. The tests employed included Forms A and B of the short vocabulary test originally developed by E. L. Thorndike for the IER Intelligence Scale CAVD, the

Concept Mastery Test (Terman & Oden, 1959) devised to measure the gifted group in adulthood, and the verbal subtests (excluding Vocabulary) of the Wechsler Adult Intelligence Scale. The sample was preselected on intellectual ability since the company characteristically uses tests in the screening of applicants. All of the men had had a high school education at the minimum. Eighty-eight percent had attended college and 68% had graduated. Within the sample the most frequent occupation was that of salesman. However, approximately a quarter of the men were at the management level when tested. None had a Wechsler Verbal IQ below 100.

The intercorrelations are presented in Table 1. Forms A and B of the short vocabulary test have been combined to provide an index of the effectiveness of a longer (40 item) vocabulary measure. Form A was employed in the *Intelligence in the United States* studies. The WAIS Verbal score was computed by combining scaled scores for the Information, Comprehension, Arithmetic, Similarities, and Digit Span subtests.

Table 1 also contains information on the mean scores and standard deviations for each of the tests. The evidence for selection on intelligence or its correlates is marked. On Form A, for which we have definite information, the mean score for the population as a whole (excluding those aged nine and below) is 10.77. The standard deviation is 3.46.

The difference in means between Forms A and B is reliable ($t = 5.50$, $p < .01$). This finding has been checked on a sample of 64 tabulating machine operators who also took both forms of the test. In this latter group the means were 11.59 and 13.45, respectively, ($t = 4.95, p < .01$). Thorndike reports that Form A appears to be somewhat easier at the lower levels and harder at the higher levels than Form B. The results presented here are consistent with such a conclusion.

In view of the known restriction of range in the present sample, it seemed advisable to estimate the size of the correlations that might be expected in the population as a whole. Such figures are more suitable for evaluating the validity of the *Intelligence in the United States* studies which utilized a sample representative of the total population. They are also more suitable for making comparisons with the findings obtained in the studies employing longer vocabulary tests,

TABLE 1

OBTAINED CORRELATIONS FOR A SELECTED GROUP OF SALES EMPLOYEES ($N = 108$)

Test	*MS*	*SD*	Vocabulary Test		Concept Mastery	WAIS Verbal
			Form A	Form B		
Vocabulary Test Forms A & B (40 items)	27.64	3.77	.89	.89	.73	.56
Vocabulary Test Form A (20 items)	13.03	2.11		.59	.64	.47
Vocabulary Test Form B (20 items)	14.61	2.11			.67	.54
Concept Mastery Test	61.50	26.60				.54
WAIS Verbal	67.22	6.34				

which characteristically were based on samples of hospital patients, prison inmates, service inductees, unselected school children, or the adult population.

The estimates in Table 2 which involve the Concept Mastery Test and WAIS were computed using the following formula taken from Gulliksen (1950):

$$R_{XY} = \sqrt{1 - (1 - r_{xy}^2)s_y^2/S_Y^2}$$

In this formula, the standard deviation of the population scores for both forms of the vocabulary test (S_y) was taken as 3.46. In all probability the resulting estimates, which average in the high .80s and tend generally to exceed by a small amount the estimate of reliability in the population, are somewhat inflated. Apparently the standard deviations obtained in the selected groups for the two forms of the vocabulary test (s_y) are underestimated due to ceiling effects. Inspection of the distributions suggests, however, that the error introduced in this manner is not sizable. It seems almost certain that the population figures would not drop below .75.

The correlation in Table 2 between Form A and Form B is an estimate of reliability in the population as a whole and was derived using Gulliksen's Equation 5 (1950, p. 111). This assumes the two are parallel forms, which is apparently, not strictly speaking, the case. Nevertheless, the figure is consistent with those previously reported. A similar estimate was derived from the tabulating machine operator sample. The corrected reliability coefficient there was .82. The estimates involving Form B are based on the assumption that the standard deviation in the population is the same as for Form A. This assumption in all probability yields excessively conservative estimates of correlation

TABLE 2

ESTIMATED CORRELATIONS BETWEEN TESTS IN THE GENERAL POPULATION

	Vocabulary Test Form B (20 item)	Concept Mastery Test	WAIS Verbal
Vocabulary Test Form A (20 items)	.85	.88	.84
Vocabulary Test Form B (20 items)		.89	.86

size. Lorge (1957) reports a standard deviation of 3.4 for Form A and 4.0 for Form B among United States Army enlisted men.

The above reference also contains information on correlations between the short vocabulary tests and the Army General Classification Test for two representative samples of enlisted men. The samples numbered 868 for Form A and 883 for Form B. The coefficients were .70 and .77, respectively.

CONCLUSION

Taken together the Lorge data and the present findings offer rather impressive support to those who have argued for the validity of short vocabulary tests as measures of intelligence. The 20-item tests yield correlations with the more general measures of intelligence which are comparable to those found between the general measures themselves. In the population as a whole the correlation between the short tests and tests of general intelligence appears to be at least .75. This figure compares favorably with that of .83 found for the longer and presumably somewhat more reliable vocabulary tests on which data have been reported previously.

It is hoped that the publication of these findings will result in a greater use of short vocabulary tests in connection with research studies where

practical considerations have heretofore precluded the investigation of relationships between experimental variables and intelligence. Prior assumptions regarding the impossibility of achieving adequate reliability and validity in such short measures were apparently incorrect. There would seem to be little basis for criticizing any study on the score of inadequate measurement techniques, purely because a short verbal measure of intelligence was employed.

REFERENCES

GULLIKSEN, H. *Theory of mental tests.* New York: Wiley, 1950.

HAGEN, ELIZABETH P., & THORNDIKE, R. L. Normative test data for adult males obtained by house-to-house testing. *J. educ. Psychol.*, 1955, **46**, 207–216.

LORGE, I. *Vocabulary Test G-T: Directions and norms.* New York: Columbia University, 1957. (Mimeo)

MINER, J. B. *Intelligence in the United States.* New York: Springer, 1957.

TERMAN, L. M., & ODEN, MELITA H. *The gifted group at mid-life.* Stanford: Stanford Univer. Press, 1959.

THORNDIKE, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26**, 128–135

THORNDIKE, R. L., & GALLUP, G. H. Verbal intelligence of the American adult. *J. gen. Psychol.*, 1944, **30**, 75–85.

(Received October 12, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 161-165

VISUAL PERCEPTION OF BINARY PATTERNS BY PRESCHOOL CHILDREN AND BY SCHOOL CHILDREN[1]

DOROTHY W. DYER AND E. RAE HARCUM[2]
College of William and Mary

This study attempts to relate the appearance of a primacy effect in the perceptual span to progress in learning to read. The primacy effect, as usually defined, is the phenomenon that initial elements of a sequential task produce fewer errors in stimulus reproduction than elements occurring later in the sequence. In the determination of perceptual spans, the superior reproduction of elements at the left within the visual pattern has often been attributed to the development of a learned set or tendency to perceive first those elements at the left within the pattern (Anderson, 1946). Therefore, these elements on the left would be favored by a primacy effect. Since this implied left-to-right viewing sequence is appropriate for reading English, the thesis of the present paper is that it has been developed through practice in reading English. By the same argument, more accurate reproduction for stimuli appearing in the right visual field would suggest a right-to-left viewing sequence and, consequently, lesser development in reading skill.

Learning or experience has been found to affect performance on tasks which require the perception of spatial relationships. For example, the tendency to make perceptual reversals decreases as progress is made in learning to read. Smith's (1950) "retarded" readers made approximately 50% more reversal errors than did a group of "advanced" readers. Similarly, inferior and superior readers exhibited differences in the relative number of errors in reproducing tachistoscopic letter-stimuli on the two sides of fixation (Crosland, 1939). Although the superior readers made fewer errors overall, for the stimuli in the right visual field the inferior readers exhibited greater accuracy. Nevertheless, the averages for both superior and inferior readers revealed greater accuracy for elements in the left field of vision. When Anderson (1946) presented nonsense English and Hebrew words to bilingual observers, the observers recognized more English letters to the left of fixation, and more Hebrew letters to the right of fixation. To account for this, Anderson suggested a learned predisposition to perceive first the left-hand letters in English words, and the right-hand letters in Hebrew words—which are, of course, read from right to left.

Heron (1957), also, found fewer errors in recognizing English letters

[1] This article is based on a dissertation submitted by the senior author to the Department of Psychology of the College of William and Mary in partial fulfillment of the requirements for the degree of Master of Arts (Dyer, 1960). This investigation was supported by PHS Research grants (B-1971 and B-2661) from the Institute of Neurological Diseases and Blindness, United States Public Health Service.

[2] Our thanks are due to Anne Smith, Head of the Happy Hours School, Williamsburg, Virginia, and to Jeanne Etheridge, Principal of the Matthew Whaley Elementary School, Williamsburg, Virginia, for their kind cooperation and assistance in the conduct of this research. The authors also wish to thank S. B. Williams and Gertrude Johnson for helpful suggestions and critical reading of the manuscript.

in the left visual field when groups of letters were presented on both sides of fixation. According to Heron, a process involving a spatial-temporal analysis of the stimulus trace resulting from even a single glance favors the letter-stimuli appearing to the left of fixation. This post-exposure process, as Heron calls this sequential analysis of the afterimage, illustrates how sequential viewing habits transferred from the reading situation might affect the perception of tachistoscopically presented material, as well as the perception of material requiring successive eye movements.

Tachistoscopically presented nonalphabetical stimuli produce results similar to those obtained with groups of letters. Harcum (1957), using linear patterns of filled and unfilled circles, found generally superior reproduction for those elements in the left visual field. This result suggests that patterns of meaningless forms may be used to study the development of a so-called normal viewing sequence. The use of such stimuli would allow the investigation of the basic perceptual mechanism without contamination by the meaningfulness, dependencies, or other discriminating characteristics in verbal stimuli.

The present study investigates the development of perceptual organization for visual stimuli using meaningless stimulus-elements simultaneously presented on both sides of fixation. Two hypotheses are tested: observers with no formal reading training reveal little or no difference in reproduction accuracy for stimuli in either visual field; and observers who have had reading training exhibit superior reproduction of stimuli in the left visual field, because they have learned to attend first to stimuli in that hemifield when learning to read.

Method

Sixty-six children with normal vision served as observers. Nursery school (N = 13) and kindergarten (N = 18) observers were enrolled in the same private school. All first grade (N = 18) and second grade (N = 17) observers had previously attended that same private school and were pupils at a public elementary school. All children returning a signed parental permission slip were tested, except for the kindergarten children of which only the first 18 were tested.

The apparatus was a Dodge-type tachistoscope. The illumination on the stimulus field and the fixation field, provided by two incandescent light sources, produced a field luminance of about 2 foot-Lamberts. The observers appeared to experience no difficulty in detecting the individual stimulus elements.

Each stimulus template consisted of six horizontally arranged ellipses. The ellipses were .360 inches (31.7′) in height, .315 inches (27.8′) in width, and were spaced .400 inches (35.3′) apart. Stroke width of the ellipses was about .035 inches (3.1′). The fixation cross on the fixation field registered with the center of the template at a distance of 26 inches from the observer's eyes. The target field was 4.5 inches (9.9°) square.

Individual patterns were made by blackening one element of the template on the left and one element on the right of the fixation cross. Each of the six element-positions was filled an equal number of times.

Each observer, viewing binocularly, first reproduced three practice patterns. Each of nine patterns was then observed (in different sequences) with binocular, left-eye, and right-eye viewing. Each observer was assigned one of the six possible viewing orders of the three viewing conditions. An attempt was made to reduce fixational errors by requiring the observer to report the color (which was periodically changed) of the fixation cross prior to exposure of the pattern. Immediately after each .15-second exposure of the pattern the observer pointed out on a blank template before him the elements that corresponded with the ones he had observed as blackened in the stimulus. An error was recorded when an observer incorrectly indicated that an open element was filled, or when he failed to indicate a blackened element.

The intertrial interval was determined by the pace at which the observer chose to

work. Testing time for each observer was usually 15–20 minutes.

A Snellen Chart was used to determine the visual acuity of the observer, and a manoptoscope was used to check his handedness and eyedness. Reading readiness of the kindergarten observers was indicated by performance on Sections 1–4 of the Metropolitan Readiness Tests. Teachers' ratings were obtained to estimate the reading proficiency of the first and second grade observers. Circumstances prevented giving the Metropolitan test to the nursery school children.

Results

The data from the binocular viewing condition for the four observer groups are shown in Figure 1, in which the means of errors per exposure are plotted as a function of the element-position from the left end of the pattern. The data for the two monocular viewing conditions are not shown because they almost exactly duplicate the binocular data.

As would be expected, there is a decrease in errors associated with the higher educational (maturational) levels.

Since the shapes of the error-curves for the nursery school and kindergarten observers appear basically the same, as do the curves for the first and second grade observers, the observers are categorized for purposes of statistical analysis of the hemifield differences into two general groups of preschool vs. school observers. A small difference in errors favoring the right visual field appears for the preschool observers. When the frequency of observers who exhibit superior performance to the right of fixation are compared to the expected frequency under the null hypothesis, the obtained difference is not significant by the binomial test ($p > .42$). However, the school children exhibit a consistent difference favoring the stimuli in the right visual field. This difference,

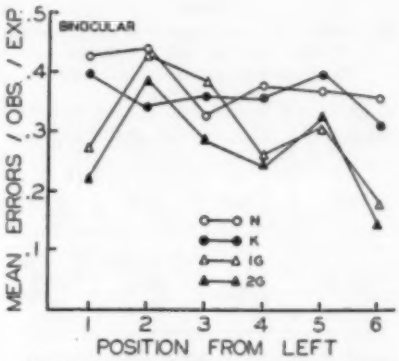


Fig. 1. Means of errors per element-position for each observer group under the binocular viewing condition.

tested as before by the binomial test, is significant ($p < .012$). The marked change in the results of preschool and school children is especially impressive when one considers the wide range of developmental stages represented within each educational level. For example, one child in kindergarten was highly proficient in reading, and his error-curves were quite similar to those of the first and second graders.

The consistent and regular curves in Figure 1 represent the averaging out of individual differences in the relative numbers of errors left and right of fixation. However, these differences in performance between the left and right fields came closer to canceling out for the preschool children than for the school children.

The data of the school children exhibit a decrease in errors at the most eccentric element-positions and for the elements nearest fixation. These curves are similar to those frequently seen in the data of adult observers on a similar task, except for the presence of fewer errors for elements to the right of fixation.

Table 1 presents the comparison of the measures of handedness and eyed-

TABLE 1

RELATION OF EYEDNESS AND HANDEDNESS TO TOTAL ERRORS PER OBSERVER AND TO PERCENTAGE ERRORS LEFT OF FIXATION

Handedness	Measure	Left-eyed	Right-eyed
		(N = 8)	(N = 5)
Left	Total	49.0	58.6
	% L	55	53
		(N = 20)	(N = 33)
Right	Total	52.8	53.2
	% L	56	53

ness with the total number of errors made by each observer, and the percentage of these errors that occurred for elements to the left of fixation. Neither the total number of errors made by an observer, nor the percentage of his errors which were made for elements left of fixation, are significantly related to laterality preferences.

Table 2 shows a comparison of ratings of higher vs. lower reading proficiency and the occurrence of higher vs. lower relative number of errors to the left of fixation. The frequencies in this two-way classification are based only on the data of the first and second grade observers. These two classifications are not significantly related ($\chi^2 = 2.40$; $p > .05$), although the difference is in the direction of a positive relation between relatively more errors left of fixation and teachers' ratings of above average reading proficiency.

DISCUSSION

The present study provides evidence in support of the first hypothesis —that the preschool observers do not manifest differences in accuracy of performance between the left and right hemifields. However, the second hypothesis, which specifically predicts the development of superiority for elements in the left visual field coincident with the onset of instruction in reading, is not confirmed. Fewer errors were made in reproducing stimuli in the right visual field, rather than in the left visual field. Since this significant difference in the relative number of errors appeared rather abruptly with the start of schooling, and since it was not found in the data of either of the two preschool groups, it would seem more reasonably due to instruction in reading rather than to maturation.

Several mechanisms may be proposed for relating instruction in reading to superior reproduction for stimuli in the right visual field. Children who have had instruction in reading may tend to look toward the right end of the visual pattern (i.e., fixate incorrectly) in this task. But this does not seem likely. Similarly, the children may have acquired a right-to-left viewing sequence in the early stages of learning to read, but this also seems unlikely. On the other hand, the children may have developed a left-to-right scanning sequence as predicted, but, for observers of this developmental level, recency effects may be stronger than primacy effects. This latter mechanism seems to be the most plausible in view of the small, although not statistically significant, positive relationship between ratings of reading proficiency and greater

TABLE 2

RELATION OF READING RATINGS TO THE PERCENTAGE ERRORS TO THE LEFT OF FIXATION FOR FIRST AND SECOND GRADE OBSERVERS

% Left	Average reader or below	Above average reader
Above 50%	9	13
50% or below	6	7

numbers of errors to the left of fixation.

Summary

Very simple binary patterns were observed by nursery school, kindergarten, first grade, and second grade school children with left-eye, right-eye, and binocular viewing. It was hypothesized, on the basis of data previously obtained from adult observers, that the relative accuracy for reproducing elements to the left of fixation would increase with school training in the left-to-right reading sequence of English.

The preschool observers exhibited nonsignificantly greater accuracy for stimuli in the right hemifield. Contrary to the hypothesis, significant right hemifield superiority was observed in the school children. Apparently, perceptual mechanisms of children at the age at which they start school do not wholly duplicate adult mechanisms.

No important effect of viewing-eye condition, or of eyedness, handedness, or "conflicting" eyedness-handedness of the observers was found.

REFERENCES

Anderson, I. H. The effect of letter-position on range of apprehension scores, with special reference to reading disability. *U. Mich. Sch. Educ. Bull.*, 1946, **18**, 37–40.

Crosland, H. R. Superior elementary-school readers contrasted with inferior readers in letter-position, "range of attention," scores. *J. educ. Res.*, 1939, **32**, 410–427.

Dyer, Dorothy W. Perception of binary visual patterns by pre-school children and by school children. Unpublished master's dissertation, College of William and Mary, 1960.

Harcum, E. R. Three inferred factors in the visual recognition of binary targets. In J. W. Wulfeck & J. H. Taylor (Eds.), *Form discrimination as related to military problems.* (NAS-NRC Publ. 561) Washington, D. C.: NAS-NRC, 1957. Pp. 32–37.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

Smith, Linda C. A study of laterality characteristics of retarded readers and reading achievers. *J. exp. Educ.*, 1950, **18**, 321–329.

(Received November 4, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 166–172

FIXED SEQUENCE VERSUS BRANCHING AUTOINSTRUCTIONAL METHODS

HARRY F. SILBERMAN, RALPH J. MELARAGNO, JOHN E. COULSON, AND DONALD ESTAVAN

System Development Corporation

Two basic autoinstructional methods may be identified. The one most frequently used displays a fixed sequence of teaching items, requiring all subjects to traverse all items (Skinner, 1958). An alternative method has been suggested (Coulson & Silberman, 1960; Crowder, 1960) which uses a variable sequence of items, allowing each student to branch to remedial materials when needed.

With a fixed-sequence teaching machine, adaptation to individual differences is provided only by the variable rate at which students move through the sequence. Variable-sequence teaching machines provide for individual differences by using a variety of criteria, such as error rate and error type, for branching to additional remedial items and for skipping over redundant items. Each student generates his own unique sequence of items to effect greater learning efficiency. The basis for the design of an autoinstructional method in which the sequence of instructional material is tailored to the needs of each student is the notion that there is no single best item sequence nor one best method of presenting the materials. It is assumed that different students require different items with varying amounts of review at different times during training.

Coulson and Silberman (1960) attempted to evaluate the effectiveness of branching versus fixed sequence instructional methods. Using a manually operated machine they compared a fixed item sequence with a variable sequence in which some items were skipped by students who made no errors. The findings showed that students trained with the very simple branching procedure used in that study learned as well as students trained without branching, and also learned in a shorter time. The amount of instructional material was not controlled, since subjects in the branching group had skipped certain items seen by the fixed-sequence group. In such a study, obtained differences can be attributed to the difference in items presented as well as to the effects of adjusting the instructional sequences to the needs of individual students.

In the Coulson and Silberman (1960) study, the wishes of the student were not considered in determining whether he was branched back to preceding items or skipped ahead to subsequent material. Nor was error rate over a sequence of items used as a branching criterion. The branching depended entirely on the errors made on individual items.

The present study was concerned with two methods of branching. In Experiment I, a procedure that permitted the student to branch at his own option was evaluated, while in Experiment II the use of errors as a branching criterion was attempted. The hypothesis tested in Experiment I was that optional-branching sequences would be superior to a fixed sequence; the hypothesis in Experiment II was that an error-generated branching sequence would be superior

to a fixed sequence when identical items are used in each sequence.

Experiment I

Two hypotheses relating to optional branching were tested. The first was that a fixed-sequence program would be less effective than the same items cast as statements in textbook format through which the student could branch at his own option. The ease of reviewing and the facility for skipping ahead with the textbook format were the bases of this hypothesis. The second hypothesis was that performance on a program which allowed the student to review previous material, one item at a time, would also exceed performance on a fixed-sequence program.

Procedure

Students from five high schools in Los Angeles County served as paid subjects. Subjects were obtained from junior and senior classes, and were pretested with the Henmon-Nelson Test of Mental Ability (Form A, Grades 9–12). Random assignment to three treatment groups resulted in 7 males and 10 females being assigned to two groups, and 6 males and 11 females being assigned to the third group, providing 17 subjects in each group.

Sixty-one multiple-choice teaching items on some topics in logic were prepared and given a preliminary trial in a number of secondary schools. Items were revised, typed on 5 × 8 cards, and mounted in loose-leaf binders for the experiment. Topics covered included simple and compound statements, connectives and their combinations, and arguments.

Three instructional methods were used. The first group, designated "fixed sequence," (Henmon-Nelson Mean = 114.1) received the 61 teaching items in a fixed sequence. Subjects in this group were instructed to give a covert response to each item in turn, and to compare their answer with the correct answer, which was inscribed on the back of the card. Subjects were further instructed not to turn back to preceding items. The second group, designated "back branching," (Henmon-Nelson Mean = 113.1) received the same 61 teaching items but was permitted to review preceding items by backing up one item-card at a time. Subjects in this group were encouraged to review earlier material and were instructed not to skip any items, so that they would view each item at least once. The third group, designated "textbook," (Henmon-Nelson Mean = 111.3) received a typewritten text produced by rewriting the original items. Each item was converted from a question into a statement by incorporating the correct answers, eliminating distractors and filling in blanks. These statements were then organized into paragraphs. Subjects in the third group were instructed to study the material in whatever manner they chose. All subjects were observed through one-way windows to insure that they were following directions and, additionally, to check on whether subjects in the back-branching group exercised the branching capability.

Mastery of the topics covered in the teaching session was tested with 24 free-response and 24 multiple-choice test questions. Half of the questions were similar to the actual training materials and half consisted of situations which required application or transfer of learning. Split-half reliability of the posttest was .93 (after Spearman-Brown correction). The test was administered immediately after the training session.

Results

Comparison of the criterion scores (Table 1) for the textbook and the fixed-sequence groups confirms the hypothesis that the textbook method, which permits the student to branch at his own option, is superior. The second hypothesis, that the back-branching method would be superior to the fixed-sequence method, was not confirmed.

TABLE 1

Posttest Scores and Training Time

Treatment	*N*	Criterion Posttest		Time (in minutes)	
		M	*SD*	*M*	*SD*
Fixed Sequence	17	28.4	8.7	32.8	10.7
Back Branching	17	30.8	9.7	31.8	9.6
Textbook	17	33.7	6.4	31.5	11.1

TABLE 2

CORRELATIONS OF APTITUDE, TRAINING TIME, AND POSTTEST SCORE

	Fixed Sequence	Back Branching	Textbook
Aptitude vs. Score	.45	.39	.67**
Aptitude vs. Time	−.52*	−.45	−.31
Time vs. Score	−.14	.03	−.04

* $p < .05$.
** $p < .01$.

For the first hypothesis, that the textbook method would be superior to the fixed-sequence method, $t = 1.79$, $p < .05$. For the second hypothesis, that the back-branching method would be superior to the fixed-sequence method, $t = .80$, $p > .10$.

Correlations of Henmon-Nelson aptitude scores, criterion posttest scores, and training times are given in Table 2.

EXPERIMENT II

A branching program, using error rate as a branching decision criterion was compared with a fixed-sequence program with instructional materials equated for the two groups. It was hypothesized that performance of subjects receiving a branching sequence would be superior, with respect to scores on a criterion posttest, to that of students receiving a fixed sequence. This was predicted on the basis of the teaching machine's responsiveness to the student's performance. By providing extra remedial items to students who have difficulty, the machine tends to minimize the number of items seen by the brighter student and to minimize the number of errors on training items made by the low aptitude student. The overall effect is intended to increase the amount of learning in a given time for each individual.

Method

Apparatus. A high speed, general purpose digital computer (Bendix G-15), equipped with a 600-slide, random-access slide projector, and an electric typewriter, was used as a teaching machine. Each slide contained one multiple-choice instructional item. Subject used the typewriter to insert his answer and receive knowledge of results.

The computer was programed to select a variable sequence of slides from a projector in response to different errors made by subject. If subject was exceeding the number of errors that had been prespecified for a particular topic, he would be branched to a set of less difficult items. If he did poorly on this set he would be shunted to still easier sets on the same topic before moving on to the next topic. Some of the multiple-choice items contained distractors the choice of which indicated a particular misunderstanding. Subjects who selected these options were branched immediately to remedial slides designed to clear up that point.

Instructional Materials. Four hundred and eleven multiple-choice items on logic (Silberman, Coulson, Melaragno, Estavan, & Englund, 1960) were prepared and given a preliminary trial in a number of secondary schools. Items were revised and put onto 35-mm. slides for use in the random-access projector.

Topics covered by these items included simple and compound statements, logical connectives, and six argument forms. Instructional items were arranged into difficulty levels. Items at one level were written so as to contain a maximum amount of information, and were the most difficult items. The items at the second level were essentially in the same format as those of the first level except that pertinent information was generally distributed over more slides than at the first level. The same was true of the third and fourth levels; here the items were designed to be relatively easy to answer, in that the step between items was kept small and the prompts were increased.

Subjects. Thirty-six students (21 males and 15 females) were selected from four high schools in Los Angeles County. Subjects had either just graduated or were in their last year of high school. Subjects were pretested with the Henmon-Nelson Test of Mental Ability (Form A, Grades 9–12) and were randomly assigned to two groups. Subjects were naive with respect to the subject matter (logic) used in the lesson.

Procedure. One treatment group was des-

ignated the branching group and the other the fixed-sequence or nonbranching group. Members of the branching group received sequences of items determined by the errors that were made during the teaching session. The computer kept cumulative error records and also a record of errors on individual items. Throughout the session the machine selected an appropriate sequence of instructional items for each student based upon error criteria. In this way, the bright student was given a minimum of material, while the slower student was given extra, remedial material on topics in which errors were committed.

Each member of the fixed-sequence group was paired at random with one member of the branching group. The unique sequence of items which each subject in the branching group generated was presented to his mate in the fixed-sequence group. Thus, pairs of subjects in the fixed-sequence and the branching groups received *identical* item sequences. The difference between the two groups was that the machine was responsive to the errors made by subjects in the branching group but was not responsive to errors made by subjects in the nonbranching group.

If the brighter member of a pair of subjects was in the branching group, the resulting sequence was expected to be too brief to provide effective instruction for his mate in the fixed-sequence group. If the brighter member of a pair of subjects was in the fixed-sequence group, the sequence which he received was expected to be longer than necessary for him, resulting in wasted time and boredom.

Knowledge of results, which consisted of a typewriter printout of the number of the correct answer, was the same for both groups. One subject was trained at a time.

Subjects were given an explanation of the operation of the equipment and permitted to familiarize themselves with the machine on the first two introductory items. In the middle of the session, subjects were given a short break. Immediately after training was completed, the criterion posttest was administered to each subject. No time restrictions were placed on subjects either during the training period or during the posttest.

Criterion Test. Mastery of the topics covered by the training items was tested with 51 multiple-choice test items and 44 free-response test items. Half of the test items were quite similar to the actual training materials and the other half required application or transfer of learning to situations not encountered in the training series. The split-half reliability of the test was .97 (after Spearman-Brown correction).

TABLE 3

APTITUDE, TRAINING, AND POSTTRAINING PERFORMANCE

		Treatment Groups[a]	
		Branching	Fixed Sequence
Number of Instructional Items	*M*	156.00	156.00
	Range	69–248	69–248
Training Errors	*M*	24.66	21.94
	SD	17.30	13.57
Training Time (minutes)	*M*	78.83	81.61
	SD	27.34	29.21
Henmon-Nelson IQ's	*M*	123.61	122.61
	SD	19.16	17.89
Posttraining Criterion Scores	*M*	66.33	64.33
	SD	15.93	16.42

[a] $N = 18$ for each treatment group.

Results

Table 3 gives data on the number of items traversed by subjects, the number of errors made on instructional items, and the training time. The mean time required for each item was 17 seconds. Table 3 also contains information on the Henmon-Nelson aptitude pretest and the criterion posttest.

Covariance analysis of criterion posttest scores, controlling training time and aptitude, yielded no significant difference between the branching and the fixed-sequence groups.

Intercorrelations of aptitude, criterion score, training time, and instructional errors are given in Table 4.

DISCUSSION

While both experiments were concerned with branching, they are not

TABLE 4

INTERCORRELATIONS OF APTITUDE, CRITERION SCORES, TRAINING TIME, AND INSTRUCTIONAL ERRORS

	Branching	Fixed Sequence
Aptitude vs. Score	.763*	.845*
Aptitude vs. Errors	−.677*	−.769*
Aptitude vs. Time	−.833*	−.407
Time vs. Score	−.682*	−.179
Errors vs. Score	−.904*	−.846*
Errors vs. Time	.793*	.472

* $p < .01$.

directly comparable in that different apparatus and procedures were used.

Overall, the results of Experiment I indicate that the relationship between flexibility of materials and rate of learning favors the flexible materials over the fixed sequence. Certain implications for further exploration are evident when this study is compared with those by Holland (1960) and Cook and Spitzer (1960), which contain features similar to it.

With a semester course in introductory psychology as his subject matter, Holland also used three instructional groups. Members of the first group were trained in the conventional manner, with confirmation of the correct answer following subject's overt response. In his second group, subjects received no confirmation following their response.

Holland found no difference in performance between the confirmation program and the same program without the confirmation. Both of these programs were superior to a no-response, control program which only required subjects to read complete statements. Holland interpreted his results as lending support for the principle that an overt response has to occur for efficient learning. Insofar as Holland's control group can be thought of as using a prompting procedure, his results are in support of confirmation as a better teaching method.

Cook and Spitzer (1960) reported results of a study using a short list of paired-associates which seem to conflict with Holland's findings. A 2 × 2 design was employed in which the experimental conditions were prompting and confirmation, with overt response or no response used in each condition. Cook and Spitzer concluded that overt responses interfered with learning and that the prompting procedure was superior to confirmation because of the enforced delay between stimulus and response terms in the confirmation method.

The superiority of the textbook group over the fixed-sequence group in Experiment I lends support to the prompting side of the prompting-confirming controversy, when no overt response is required, since the textbook condition corresponds to the prompting procedure while the fixed-sequence condition used a confirmation procedure.

Since the three studies involved materials which were dissimilar in length and quality, it remains for further research to compare the influences of response requirement, S-R delay, and optional selection of material.

Two possible conclusions can be obtained from Experiment II. The first is that a teaching machine which branches on the basis of errors is not significantly superior to a machine which displays a fixed sequence of items. An alternative explanation is that the error criteria used for branching in this experiment, and the remedial items to which a student who was having difficulty was branched, were inefficient and acted to obscure any real differences between the two methods. Certain of the experimental results lend support to the second conclusion.

Analysis of the criterion tests shows that low aptitude subjects in both groups failed to learn much of the material. While this had been predicted for the fixed-sequence group, it also had been expected that members of the branching group, who would supposedly receive a sequence uniquely tailored to their individual needs, would receive the training necessary to insure their learning the material. This was the case for only two (of seven) subjects with low aptitudes in the branching group. Both of these subjects were frequently branched to remedial items, and both scored above their group's mean on the criterion test. The five remaining low aptitude subjects in the branching group failed to score at or near their group's mean. All six subjects in the fixed-sequence group with low aptitudes fell well below their group's mean.

Three factors which may account for the ineffectiveness of the remedial items for low aptitude subjects are: the limited variety of these items, the inability of subjects to branch out of the remedial series when their difficulties had been resolved, and the inappropriateness of the particular measure of individual differences (errors) used as the criterion for branching decisions.

Since the instructional items in the remedial branches were very similar to those in the main series, except that the subject matter had been divided into small steps, the mode of instruction in Experiment II was of limited versatility. That is, no attempt was made to alter the teaching method for a given subject who was having difficulty. It is postulated that the low aptitude subjects who scored well actually received remedial material which assisted them, while the low aptitude subjects who did poorly were not assisted because the machine simply was not flexible enough to find a set of items adequate for them.

When a low aptitude subject was branched to remedial material, no attempt was made to evaluate the subject's correct responses in order to branch him back to the main set of items when he demonstrated proficiency. Instead, subject was obliged to remain in the remedial branch until he had negotiated all of the items in that branch. For many of the low aptitude subjects, this meant that they were forced to see more items than necessary.

In Experiment II it was assumed that the dimensions of individual differences measured by errors on instructional items interacted with the effectiveness of the item sequence presented. The machine was, consequently, programed to treat subjects differentially (i.e., giving different item sequences) with respect to the dimension of error rate. The results show that the particular method of providing for individual differences used in this experiment was not a sufficient condition for effective learning of the logic lesson. Perhaps the common principle of teaching that one should "provide for individual differences" needs to be qualified with the specific conditions for its accomplishment. It may be conjectured that some measures such as response latency or subject's self-evaluation are more appropriate than error rate, and that the computer should have considered these behavior measures for its branching decisions instead of, or in addition to, errors.

Summary

The experiments investigated the effectiveness of two methods of branching, as measured by posttraining criterion tests. In Experiment I, three groups of 17 subjects were used to test

two hypotheses concerning optional branching. A fixed-sequence group received items in fixed order; a back-branching group receiving the same items as the first group, was permitted to back up one item at a time to review earlier items; a third group received the same items cast in statement form and organized into paragraphs permitting subjects to choose material at their own option. A significant difference on a posttest in favor of the third group was obtained when the first and third groups were compared.

In Experiment II, a computer-controlled teaching machine was used to evaluate the effectiveness of adapting sequences of teaching items on logic.

Members of a branching group received sequences of items determined by the errors that were made during instruction. Each member of a fixed sequence group was paired at random with one member of the branching group. The particular sequence of items which each subject in the branching group received was presented to his mate in the fixed-sequence group. Covariance analysis of criterion scores using aptitude and training time as control variables yielded no significant difference between the branching and fixed-sequence conditions. A discussion of the limitations of the study and of possibilities for further research on this problem was presented.

REFERENCES

Cook, J. O., & Spitzer, M. E. Supplementary report: Prompting versus confirmation in paired associate learning. *J. exp. Psychol.*, 1960, **59**, 275–276.

Coulson, J. E., & Silberman, H. F. Effects of three variables in a teaching machine. *J. educ. Psychol.*, 1960, **51**, 135–143.

Crowder, N. A. Automatic tutoring by intrinsic programming. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington: National Education Association, 1960.

Holland, J. G. Program design and use. Paper read at American Psychological Association, Chicago, September 1960.

Silberman, H. F., Coulson, J. E., Melaragno, R. J., Estavan, D. P., & Englund, D. E. *Teaching machine program: Logic*. (TM-534) Santa Monica, Calif.: System Development Corporation, 1960.

Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

(Received November 7, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 3, 173–178

A RE-EXAMINATION OF COLLEGE PRODUCTIVITY[1]

ALEXANDER W. ASTIN[2]

National Merit Scholarship Corporation

Several recent investigations have attempted to evaluate what part the undergraduate's educational experience plays in stimulating him to go on for the PhD degree. The first of these, the studies of Knapp and Goodrich (1952) and Knapp and Greenbaum (1953), defined the productivity of any undergraduate college as the percentage of its graduates who eventually obtain the PhD. In a subsequent study, Holland (1957) showed that an institution's productivity is related in part to the academic ability of the students who enroll; that is, the highly productive institution has a greater proportion of talented students among its undergraduates than does the less productive institution. Following this lead, Thistlethwaite (1959a, 1959b) equated colleges with respect to student ability by adjusting each college's PhD productivity rates in terms of its "talent supply" (that is, the estimated academic ability of the student body). He found correlations as high as .83 between these adjusted productivity rates and certain "environmental press" characteristics of 36 colleges. Productivity of PhD's in the natural sciences (NS) was correlated with a somewhat different set of press than was productivity in the arts, humanities, and social sciences (AHSS).

The present study attempts to extend this work by exploring the following questions:

1. Is ability the only quality of the student body which is related to a college's productivity? Are there not other systematic differences among the student bodies enrolling at various institutions which are related to PhD productivity? One such difference might be in the proportion of entering students who plan careers in either NS or AHSS. Certainly any college which is attended predominantly by students who plan careers in the natural sciences can be expected to have a relatively high proportion of graduates who go on for the PhD in NS, independent of any special effects of the college environment. Conversely, this same college would probably have a low productivity rate of AHSS PhD's. Two other characteristics of entering students which might affect the productivity of a given school are the ratio of men to women, and the proportion of students who already intend to get the PhD degree.

2. If the educational goals, major fields, and sex ratios of entering students are related to productivity, are they also responsible in some measure for the correlations found earlier between "college press" and productivity? These correlations have been interpreted as evidence for a causative role of the college on the student's motivation; e.g., "the results . . . indicate that the teacher plays an important role in motivating talented undergraduates to seek advanced degrees" (Thistlethwaite, 1959a, p. 75); "The analysis confirms once again the view that one type of college press stimulates achievement in the natural sciences, while a different type facilitates achievement in the arts,

[1] This study is a part of the research program at the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation.

[2] The author is indebted to John L. Holland, Laura Kent, and Donald L. Thistlethwaite for reviewing the manuscript.

humanities, and social sciences" (Thistlethwaite, 1959b, p. 188). However, since the information on college press came from students at each institution, it may reflect differences among student bodies which existed prior to the students' entrance into college, differences which are not a product of any unique qualities of the college environment. Before we can accurately assess the motivating function of the college itself, we must find out the extent to which these "precollege" characteristics of the student body are related to the college's productivity.

Specifically, this study examined:

1. The extent to which productivity rates of NS and AHSS PhD's (adjusted for talent supply) are a function of: (*a*) the proportion of students who, prior to entering college, intend to major in natural science; (*b*) the proportion of students who, prior to entering college, aspire to the PhD degree; and (*c*) the ratio of men to women among students enrolling in the college.

2. The extent to which relationships previously obtained between adjusted PhD productivity and college press are modified, when the effects of these additional precollege variables are removed.

Method

College Input Measures

Information on three precollege or input variables—the student's sex, probable college major, and highest degree aspired to—was drawn from a questionnaire which had been mailed to a 10% random sample of all high school students who scored above the 64th percentile on the 1959 National Merit Scholarship Qualifying Test. Of the students who returned the questionnaires, 2,884 attended one of the 36 institutions about which college press data had been obtained in the earlier studies. The number of students attending each institution ranged from 15 to 208, with a median of 72.

The following three input measures were obtained for each of the 36 colleges: the percentage of students who originally planned to major in natural science (% NS Majors), the percentage of students aspiring to the PhD degree (% PhD Aspirants), and the percentage of males among the entering students (% Males).[3]

Results

Table 1 shows the correlations between the three input measures and the adjusted productivity indexes of the 36 colleges.[4] It is evident that a significant amount of the variation in adjusted PhD productivity can still be accounted for on the basis of differences among the entering students. To obtain a productivity index which is independent of the quality of the student body, then, requires the use of corrections for variables other than just aptitude level.

The intercorrelations of the three input variables are given in Table 2. To estimate the unique contributions of these variables to productivity, their partial correlations with productivity rates were computed (Table 3).

[3] In order to make these distributions more nearly normal, all percentages were converted by means of the arcsin transformation (Guilford, 1954, pp. 574–576). Since these input measures were based on limited samples of the 1959 freshman classes at the 36 colleges, two checks were made on the representativeness of the NMSQT samples. First, % Males was correlated with the percentage of males among the entire freshman class entering each of these 36 institutions in Fall, 1958 (as reported by United States Office of Education, 1959). This coefficient was .97. Second, mean Scholastic Aptitude Test scores, which were available on the entire 1956 freshman classes enrolling at 22 of the 36 colleges, were correlated with the corresponding mean NMSQT scores of the present samples. This coefficient was .86.

[4] In computing these correlations it is necessary to assume that the relative status of the 36 colleges on the three input variables has not changed greatly over a 10- to 15-year period. In earlier reports similar assumptions were made with respect to talent supply and "college press."

Table 3 makes it clear that % Males relates to neither NS nor AHSS productivity, beyond the combined contribution of % NS Majors and % PhD Aspirants. (This would suggest, in addition, that differences in the proportions of men and women who get the PhD can be accounted for by differences in their precollege levels of aspiration and intended fields of study.) These findings prompted the decision to omit % Males as an input variable, and to deal only with % NS Majors and % PhD Aspirants.

Table 4 shows the correlations of % NS Majors and % PhD Aspirants with those Faculty and Student Press Scales which were previously reported as significantly related to adjusted productivity rates (Thistlethwaite, 1959b, pp. 187–188). Nine of the 13 college press scales are significantly ($p < .01$) related to one or both of the

TABLE 1

CORRELATIONS BETWEEN THREE INPUT VARIABLES AND THE ADJUSTED[a] PhD PRODUCTIVITY RATES OF 36 INSTITUTIONS

	Input Variables		
	% NS Majors	% PhD Aspirants	% Males
Adjusted Productivity in Natural Science (NS)	.60**	.68**	.45**
Adjusted Productivity in Arts, Humanities, and Social Science (AHSS)	−.69**	.12	−.20

[a] "Adjusted" productivity is defined as the discrepancy between a college's expected rate of producing PhD's (as predicted from its talent supply) and its actual productivity rate (Thistlethwaite, 1959b).

** $p < .01$.

TABLE 2

INTERCORRELATIONS OF THREE INPUT VARIABLES

	% PhD Aspirants	% Males
% NS Majors	.46**	.50**
% PhD Aspirants		.49**

** $p < .01$.

TABLE 3

SECOND-ORDER PARTIAL CORRELATIONS SHOWING THE INDEPENDENT CONTRIBUTION OF EACH INPUT VARIABLE TO ADJUSTED PhD PRODUCTIVITY

	Input Variables		
	% NS Majors	% PhD Aspirants	% Males
Adjusted NS Productivity	.41*	.54**	.03
Adjusted AHSS Productivity	−.83**	.66**	.00

* $p < .05$.

** $p < .01$.

TABLE 4

CORRELATIONS OF TWO INPUT VARIABLES WITH 13 COLLEGE PRESS SCALES AT 36 INSTITUTIONS

	Input Variable	
	% NS Majors	% PhD Aspirants
Student Press		
Humanism	−.39*	.36*
Breadth of interests	−.47**	.08
Reflectiveness	−.29	.45**
Participation	.12	−.41*
Aggression	.59**	.13
Scientism	.48**	.64**
Social conformity	−.36*	−.52**
Faculty Press		
Excellence of social science faculty and resources	−.71**	−.04
Flexibility of curriculum	−.56**	.03
Energy and controversiality of instruction	−.27	.31
Informality and warmth of student-faculty contacts	−.04	.55**
Closeness of supervision	.03	−.45**
Directiveness of teaching methods	−.41*	−.34*

* $p < .05$.

** $p < .01$.

TABLE 5

CONTRIBUTION OF TWO INPUT VARIABLES TO CORRELATIONS BETWEEN COLLEGE PRESS AND ADJUSTED PhD PRODUCTIVITY

	Adjusted NS Productivity		Adjusted AHSS Productivity	
	Original Correlation[a]	Partial Correlation[b]	Original Correlation[a]	Partial Correlation[b]
Student Press				
Humanism	−.07	−.25	.71**	.39*
Breadth of interests	−.27	−.26	.67**	.42*
Reflectiveness	−.05	−.38*	.67**	.40*
Participation	.00	.21	−.53**	−.42*
Aggression	.43**	.29	−.54**	−.13
Scientism	.59**	.18	.03	.33
Social conformity	−.42*	−.10	.03	−.03
Faculty Press				
Excellence of social science faculty and resources	−.42*	−.33	.83**	.55**
Flexibility of curriculum	−.31	−.24	.68**	.33
Energy and controversiality of instruction	−.13	−.35*	.58**	.35*
Informality and warmth of student-faculty contacts	.43**	.32	.40*	.14
Closeness of supervision	−.22[c]	.00	−.38*[c]	−.23
Directiveness of teaching methods	−.42**	−.16	.24	.12

[a] From Thistlethwaite (1959b, pp. 187–188).

[b] Independent of % NS Majors and % PhD Aspirants.

[c] These correlations were inadvertently interchanged in the original report.

*$p < .05$.

**$p < .01$.

input variables, and three of the other four press scales have borderline ($p < .05$) relationships with input.

In order to remove the effects of these two input variables from the correlations previously obtained between college presses and PhD productivity, second-order partial correlations were computed. Table 5 shows these correlations, along with the original correlations given in the earlier report. [5] It can be seen that all the larger correlations tended to shrink when the two input variables were partialed out. Of the 11 previously significant ($p < .01$) correlations, only one (between AHSS Productivity and "Excellence of social science faculty and resources") remained significant; five of these, all involving AHSS Productivity, were reduced to borderline significance ($p > .01, < .05$), and five to nonsignificance. In addition, all five previously borderline correlations were reduced to nonsignificance. The only reversals of the trend toward shrinkage were two previously nonsignificant correlations with NS Productivity, which became borderline as a result of the partialing procedure.

[5] Second-order partial correlations in a relatively small sample such as this should be regarded as very rough approximations. They are presented here only to show trends in the effects of student input on the original correlations.

DISCUSSION

These findings seem to warrant two conclusions. First of all, adjusting the productivity rates of undergraduate institutions solely in terms of differ-

ences in the aptitude levels of their students does not result in productivity indexes which are independent of student quality. At least two other characteristics of entering student bodies —the percentage planning to major in NS and the percentage aspiring to the PhD degree—are significantly related to adjusted productivity rates. Secondly, these input variables account for a substantial share of the relationships previously found between college environments and adjusted productivity. In short, much of the variation in PhD productivity which had been previously attributed to the motivating effects of the colleges now appears to be a function of differences among the student bodies which enroll.

An additional question posed by these analyses concerned the justification for using two productivity indexes. The high correlations of % NS Majors with both productivity rates (Table 1) suggested that breaking down productivity in this way introduces extraneous variability due to differences in the students' fields of interest. To check this notion more directly, the difference between each school's productivity rates was obtained (that is, NS rate − AHSS rate). The correlation between this difference score and % NS Majors was .84. In light of this finding, the total (unadjusted) PhD productivity rate (see Thistlethwaite, 1959b, p. 184) was computed for each college. The correlation between total productivity and % PhD Aspirants was .81. (% Males and % NS Majors were not significantly related to total productivity.) It seems apparent, then, that about two-thirds of the variation among these 36 colleges in total PhD productivity can be accounted for on the basis of a single input variable.

These findings have several implications. For one thing, they highlight one of the difficulties in attempting to assess the influence of the college on its students: differential "effects" of colleges may well turn out to be spurious, unless the relevant student input variables are controlled. For another, these results suggest that it may be necessary to reinterpret some of the results of other studies, not only more recent ones of college "press" (Thistlethwaite, 1960), but also those which utilize adjusted college productivity rates (e.g., Holland, 1960).

From the available evidence, it is not clear whether or to what extent the undergraduate institution stimulates or inhibits PhD productivity. The high correlations found between input and productivity indicate that the college plays a relatively smaller role in this sample than was previously supposed; however, we would have to study larger samples of students and colleges before we could determine precisely the relative contributions of input variables and college environments. Moreover, since this sample of colleges is far from representative,[6] due caution should be used in attempting to generalize the results to other institutions.

Summary

The PhD-producing rates of 36 colleges were subjected to further study. It was shown that a college's productivity rates are related to characteristics of its entering students other than just their academic ability; namely, the percentage planning to major in natural science and the percentage aspiring to the PhD degree. When the effects of these two input variables are partialed out, correlations previously obtained between college press and productivity rates are reduced considerably in size.

[6] These 36 colleges were selected initially because they were attended by relatively large numbers of National Merit Scholars.

REFERENCES

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HOLLAND, J. L. Undergraduate origins of American scientists. *Science*, 1957, **126**, 433–437.

HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245–254.

KNAPP, R. H., & GOODRICH, H. B. *Origins of American scientists.* Chicago: Univer. Chicago Press, 1952.

KNAPP, R. H., & GREENBAUM, J. J. *The younger American scholar: His collegiate origins.* Chicago: Univer. Chicago Press, 1953.

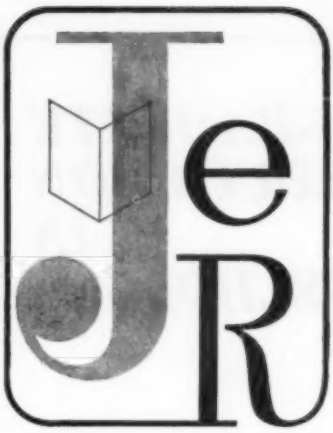
THISTLETHWAITE, D. L. College environments and the development of talent. *Science*, 1959, **130**, 71–76. (a)

THISTLETHWAITE, D. L. College press and student achievement. *J. educ. Psychol.*, 1959, **50**, 183–191. (b)

THISTLETHWAITE, D. L. College press and changes in study plans of talented students. *J. educ. Psychol.*, 1960, **51**, 222–234.

UNITED STATES OFFICE OF EDUCATION. *Opening (fall) enrollment in higher education, 1958: Institutional data.* (Circular No. 544) Washington, D. C.: United States Government Printing Office, 1959.

(Received December 16, 1960)



a personal copy on your desk . . .
the mark of the "pro" in education!

The Journal of Educational Research

For 40 years devoted to the scientific study of teaching and training. Now even bigger and better—with a new monthly digest of "What's New in Educational Research".

You can save up to $8 by subscribing now!

The Journal of Educational Research
Dembar Publications • Box 1605, Madison 3, Wisconsin

Please enter my subscription before your rates go up.

You save as much as $8 when you sign up now

☐ NEW ☐ RENEWAL ☐ 1 YEAR ☐ 2 YEARS

PRESENT RATES
*New $6
*Renewal $5
*Two Years $10
Single copy .75

RATES AFTER SEPTEMBER 1, 1961
*One Year $7.50
*Two Years $14
Single copy $1

*Plus $1 a year for postage outside U.S.

NAME _____

TITLE _____

ADDRESS _____

CITY _____ ZONE _____ STATE _____

☐ CHECK OR MONEY ORDER ENCLOSED ☐ BILL ME LATER

GRADUATE EDUCATION IN PSYCHOLOGY

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:
Anne Roe, *Chairman*,
and
John W. Gustad, Bruce V. Moore,
Sherman Ross, and Marie Skodak

Price $1.50

THE AMERICAN
PSYCHOLOGICAL ASSOCIATION
Dept. Grad
1333 Sixteenth Street, N.W.
Washington 6, D. C.